

## 7. Problem Statement

### Initial execution on terminal:

```
[acadgild@localhost ~]$ jps
3192 Jps
[acadgild@localhost ~]$ sudo service sshd start
[sudo] password for acadgild:
[acadgild@localhost ~]$ start-all.sh
This script is Deprecated. Instead use start-dfs.sh and start-yarn.sh
18/07/06 11:22:55 WARN util.NativeCodeLoader: Unable to load native-hadoop library
for your platform... using builtin-java classes where applicable
Starting namenodes on [localhost]
localhost: namenode running as process 3128. Stop it first.
localhost: starting datanode, logging to
/home/acadgild/install/hadoop/hadoop-2.6.5/logs/hadoop-acadgild-datanode-localhost.lo
caldomain.out
Starting secondary namenodes [0.0.0.0]
0.0.0.0: starting secondarynamenode, logging to
/home/acadgild/install/hadoop/hadoop-2.6.5/logs/hadoop-acadgild-secondarynamenode-
localhost.localdomain.out
18/07/06 11:23:22 WARN util.NativeCodeLoader: Unable to load native-hadoop library
for your platform... using builtin-java classes where applicable
starting yarn daemons
starting resourcemanager, logging to
/home/acadgild/install/hadoop/hadoop-2.6.5/logs/yarn-acadgild-resourcemanager-localh
ost.localdomain.out
localhost: starting nodemanager, logging to
/home/acadgild/install/hadoop/hadoop-2.6.5/logs/yarn-acadgild-nodemanager-localhost.l
ocaldomain.out
[acadgild@localhost ~]$ mr-jobhistory-daemon.sh
Usage: mr-jobhistory-daemon.sh [--config <conf-dir>] (start|stop) <mapred-command>
[acadgild@localhost ~]$ mr-jobhistory-daemon.sh start historyserver
starting historyserver, logging to
/home/acadgild/install/hadoop/hadoop-2.6.5/logs/mapred-acadgild-historyserver-localho
st.localdomain.out
[acadgild@localhost ~]$ jps
5892 DataNode
6180 ResourceManager
6023 SecondaryNameNode
6631 JobHistoryServer
```

6664 Jps  
6282 NodeManager

**Write PIG scripts for following tasks.**

### **Task 1**

**Write a program to implement wordcount using Pig.**

**Pig script:** (filename : word\_count.pig)

```
lines = LOAD '/user/acadgild/hadoopdata/pig/input/sample.txt' AS (line:chararray);
words = FOREACH lines GENERATE FLATTEN(TOKENIZE(line)) as word;
grouped = GROUP words BY word;
wordcount = FOREACH grouped GENERATE group, COUNT(words);
DUMP wordcount;
```

### **Terminal execution:**

```
[acadgild@localhost ~]$ hdfs dfs -put /home/acadgild/Desktop/sample.txt
/user/acadgild/hadoopdata/pig/input18/07/06 17:27:42 WARN util.NativeCodeLoader: Unable to load
native-hadoop library for your platform... using builtin-java classes where applicable
[acadgild@localhost ~]$ hdfs dfs -ls /user/acadgild/hadoopdata/pig/input
18/07/06 17:28:27 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your
platform... using builtin-java classes where applicable
Found 3 items
-rw-r--r--  1 acadgild supergroup    261 2018-07-06 16:56
/user/acadgild/hadoopdata/pig/input/employee_details.txt
-rw-r--r--  1 acadgild supergroup    72 2018-07-06 16:57
/user/acadgild/hadoopdata/pig/input/employee_expenses.txt
-rw-r--r--  1 acadgild supergroup    76 2018-07-06 17:27
/user/acadgild/hadoopdata/pig/input/sample.txt
[acadgild@localhost ~]$ hdfs dfs -cat /user/acadgild/hadoopdata/pig/input/sample.txt
18/07/06 17:29:40 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your
platform... using builtin-java classes where applicable
I learn bigdata
I learn hadoop
I love bigdata
it is great to learn bigdata
[acadgild@localhost ~]$

[acadgild@localhost ~]$ pig /home/acadgild/Desktop/word_count.pig
```

18/07/06 17:37:32 INFO pig.ExecTypeProvider: Trying ExecType : LOCAL  
18/07/06 17:37:32 INFO pig.ExecTypeProvider: Trying ExecType : MAPREDUCE  
18/07/06 17:37:32 INFO pig.ExecTypeProvider: Picked MAPREDUCE as the ExecType  
2018-07-06 17:37:33,054 [main] INFO org.apache.pig.Main - Apache Pig version 0.16.0 (r1746530)  
compiled Jun 01 2016, 23:10:49  
2018-07-06 17:37:33,055 [main] INFO org.apache.pig.Main - Logging error messages to:  
/home/acadgild/pig\_1530878853052.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-1.4.4/lib/slf4j-log4j12-1.7.10.jar!/org/slf4j/impl/StaticLoggerBinder.class]  
SLF4J: See [http://www.slf4j.org/codes.html#multiple\\_bindings](http://www.slf4j.org/codes.html#multiple_bindings) for an explanation.  
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]  
2018-07-06 17:37:33,988 [main] WARN org.apache.hadoop.util.NativeCodeLoader - Unable to load native-hadoop library for your platform... using builtin-java classes where applicable  
2018-07-06 17:37:34,375 [main] INFO org.apache.pig.impl.util.Utills - Default bootup file  
/home/acadgild/.pigbootup not found  
2018-07-06 17:37:34,599 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address  
2018-07-06 17:37:34,599 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS  
2018-07-06 17:37:34,603 [main] INFO org.apache.pig.backend.hadoop.executionengine.HExecutionEngine - Connecting to hadoop file system at: hdfs://localhost:8020  
2018-07-06 17:37:35,793 [main] INFO org.apache.pig.PigServer - Pig Script ID for the session: PIG-word\_count.pig-e075423c-84d8-4c50-b718-1a922019cc73  
2018-07-06 17:37:35,793 [main] WARN org.apache.pig.PigServer - ATS is disabled since yarn.timeline-service.enabled set to false  
2018-07-06 17:37:37,162 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS  
2018-07-06 17:37:37,364 [main] INFO org.apache.pig.impl.util.SpillableMemoryManager - Selected heap (Tenured Gen) of size 699072512 to monitor. collectionUsageThreshold = 489350752, usageThreshold = 489350752  
2018-07-06 17:37:37,685 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: GROUP\_BY  
2018-07-06 17:37:37,780 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS  
2018-07-06 17:37:37,787 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.  
2018-07-06 17:37:37,889 [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]}  
2018-07-06 17:37:38,185 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File concatenation threshold: 100 optimistic? false  
2018-07-06 17:37:38,251 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.util.CombinerOptimizerUtil - Choosing to move algebraic foreach to combiner  
2018-07-06 17:37:38,350 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size before optimization: 1  
2018-07-06 17:37:38,351 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size after optimization: 1  
2018-07-06 17:37:38,437 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS  
2018-07-06 17:37:38,530 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032  
2018-07-06 17:37:39,013 [main] INFO org.apache.pig.tools.pigstats.mapreduce.MRScriptState - Pig script settings are added to the job  
2018-07-06 17:37:39,033 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.reduce.markreset.buffer.percent is deprecated. Instead, use mapreduce.reduce.markreset.buffer.percent  
2018-07-06 17:37:39,033 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - mapred.job.reduce.markreset.buffer.percent is not set, set to default 0.3  
2018-07-06 17:37:39,036 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.output.compress is deprecated. Instead, use mapreduce.output.fileoutputformat.compress  
2018-07-06 17:37:39,050 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Reduce phase detected, estimating # of required reducers.  
2018-07-06 17:37:39,055 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Using reducer estimator:  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.InputSizeReducerEstimator  
2018-07-06 17:37:39,077 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.InputSizeReducerEstimator - BytesPerReducer=1000000000 maxReducers=999 totalInputFileSize=76  
2018-07-06 17:37:39,078 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Setting Parallelism to 1  
2018-07-06 17:37:39,078 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.reduce.tasks is deprecated. Instead, use mapreduce.job.reduces  
2018-07-06 17:37:39,078 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - This job cannot be converted run in-process  
2018-07-06 17:37:39,132 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.submit.replication is deprecated. Instead, use mapreduce.client.submit.file.replication  
2018-07-06 17:37:39,765 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar file:/home/acadgild/install/pig/pig-0.16.0/pig-0.16.0-core-h2.jar to DistributedCache through /tmp/temp-503356879/tmp-574728605/pig-0.16.0-core-h2.jar

2018-07-06 17:37:39,817 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar  
file:/home/acadgild/install/pig/pig-0.16.0/lib/automaton-1.11-8.jar to DistributedCache through  
/tmp/temp-503356879/tmp-145228160/automaton-1.11-8.jar  
2018-07-06 17:37:39,929 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar  
file:/home/acadgild/install/pig/pig-0.16.0/lib/antlr-runtime-3.4.jar to DistributedCache through  
/tmp/temp-503356879/tmp-1050689394/antlr-runtime-3.4.jar  
2018-07-06 17:37:40,021 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar  
file:/home/acadgild/install/pig/pig-0.16.0/lib/joda-time-2.9.3.jar to DistributedCache through  
/tmp/temp-503356879/tmp519846113/joda-time-2.9.3.jar  
2018-07-06 17:37:40,069 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Setting up  
single store job  
2018-07-06 17:37:40,092 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Key  
[pig.schematuple] is false, will not generate code.  
2018-07-06 17:37:40,092 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Starting process  
to move generated code to distributed cacche  
2018-07-06 17:37:40,093 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Setting key  
[pig.schematuple.classes] with classes to deserialize []  
2018-07-06 17:37:40,335 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 1  
map-reduce job(s) waiting for submission.  
2018-07-06 17:37:40,336 [main] INFO org.apache.hadoop.conf.Configuration.deprecation -  
mapred.job.tracker.http.address is deprecated. Instead, use mapreduce.jobtracker.http.address  
2018-07-06 17:37:40,346 [JobControl] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to  
ResourceManager at localhost/127.0.0.1:8032  
2018-07-06 17:37:40,398 [JobControl] INFO org.apache.hadoop.conf.Configuration.deprecation -  
mapred.task.id is deprecated. Instead, use mapreduce.task.attempt.id  
2018-07-06 17:37:40,409 [JobControl] INFO org.apache.hadoop.conf.Configuration.deprecation -  
fs.default.name is deprecated. Instead, use fs.defaultFS  
2018-07-06 17:37:41,012 [JobControl] WARN org.apache.hadoop.mapreduce.JobResourceUploader -  
No job jar file set. User classes may not be found. See Job or Job#setJar(String).  
2018-07-06 17:37:41,080 [JobControl] INFO org.apache.pig.builtin.PigStorage - Using  
PigTextInputFormat  
2018-07-06 17:37:41,098 [JobControl] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat  
- Total input paths to process : 1  
2018-07-06 17:37:41,098 [JobControl] INFO  
org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1  
2018-07-06 17:37:41,178 [JobControl] INFO  
org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths (combined) to  
process : 1  
2018-07-06 17:37:41,343 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter - number of  
splits:1  
2018-07-06 17:37:41,646 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter -  
Submitting tokens for job: job\_1530856936001\_0015  
2018-07-06 17:37:42,004 [JobControl] INFO org.apache.hadoop.mapred.YARNRunner - Job jar is not  
present. Not adding any jar to the list of resources.

2018-07-06 17:37:42,222 [JobControl] INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl - Submitted application application\_1530856936001\_0015

2018-07-06 17:37:42,340 [JobControl] INFO org.apache.hadoop.mapreduce.Job - The url to track the job: http://localhost:8088/proxy/application\_1530856936001\_0015/

2018-07-06 17:37:42,341 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - HadoopJobId: job\_1530856936001\_0015

2018-07-06 17:37:42,341 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Processing aliases grouped,lines,wordcount,words

2018-07-06 17:37:42,341 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - detailed locations: M: lines[1,8],words[-1,-1],wordcount[4,12],grouped[3,10] C: wordcount[4,12],grouped[3,10] R: wordcount[4,12]

2018-07-06 17:37:42,366 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 0% complete

2018-07-06 17:37:42,366 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Running jobs are [job\_1530856936001\_0015]

2018-07-06 17:38:09,890 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 50% complete

2018-07-06 17:38:09,890 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Running jobs are [job\_1530856936001\_0015]

2018-07-06 17:38:21,922 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Running jobs are [job\_1530856936001\_0015]

2018-07-06 17:38:22,942 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032

2018-07-06 17:38:22,967 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

2018-07-06 17:38:23,333 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032

2018-07-06 17:38:23,351 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

2018-07-06 17:38:23,521 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032

2018-07-06 17:38:23,542 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

2018-07-06 17:38:23,717 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 100% complete

2018-07-06 17:38:23,721 [main] INFO org.apache.pig.tools.pigstats.mapreduce.SimplePigStats - Script Statistics:

HadoopVersion	PigVersion	UserId	StartedAt	FinishedAt	Features
2.6.5	0.16.0	acadgild	2018-07-06 17:37:38	2018-07-06 17:38:23	GROUP_BY

Success!

Job Stats (time in seconds):

JobId	Maps	Reduces	MaxMapTime	MinMapTime	AvgMapTime	MedianMapTime	MaxReduceTime	MinReduceTime	AvgReduceTime	MedianReductime	Alias	Feature
job_1530856936001_0015	1	1	13	13	13	8	8	8	8	8	8	8
grouped,lines,wordcount,words			GROUP_BY,COMBINER			hdfs://localhost:8020/tmp/temp-503356879/tmp975529657,						

Input(s):

Successfully read 4 records (470 bytes) from: "/user/acadgild/hadoopdata/pig/input/sample.txt"

Output(s):

Successfully stored 9 records (109 bytes) in:

"hdfs://localhost:8020/tmp/temp-503356879/tmp975529657"

Counters:

Total records written : 9

Total bytes written : 109

Spillable Memory Manager spill count : 0

Total bags proactively spilled: 0

Total records proactively spilled: 0

Job DAG:

job\_1530856936001\_0015

2018-07-06 17:38:23,726 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032

2018-07-06 17:38:23,753 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

2018-07-06 17:38:23,992 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032

2018-07-06 17:38:24,010 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

2018-07-06 17:38:24,245 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032

2018-07-06 17:38:24,264 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

2018-07-06 17:38:24,388 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!

```
2018-07-06 17:38:24,397 [main] INFO org.apache.hadoop.conf.Configuration.deprecation -
fs.default.name is deprecated. Instead, use fs.defaultFS
2018-07-06 17:38:24,414 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key
[pig.schematuple] was not set... will not generate code.
2018-07-06 17:38:24,538 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat -
Total input paths to process : 1
2018-07-06 17:38:24,539 [main] INFO
org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(I,3)
(is,1)
(it,1)
(to,1)
(love,1)
(great,1)
(learn,3)
(hadoop,1)
(bigdata,3)
2018-07-06 17:38:24,789 [main] INFO org.apache.pig.Main - Pig script completed in 52 seconds and
253 milliseconds (52253 ms)
```

## **Task 2**

**We have employee\_details and employee\_expenses files. Use local mode while running Pig and write Pig Latin script to get below results:**

### **DataSets :**

employee\_details (EmpID,Name,Salary,DepartmentID)

[https://github.com/prateekATacadgild/DatasetsForCognizant/blob/master/employee\\_details.txt](https://github.com/prateekATacadgild/DatasetsForCognizant/blob/master/employee_details.txt)

employee\_expenses(EmpID,Expencc)

[https://github.com/prateekATacadgild/DatasetsForCognizant/blob/master/employee\\_expenses.txt](https://github.com/prateekATacadgild/DatasetsForCognizant/blob/master/employee_expenses.txt)

### **Initial Terminal execution:**

```
[acadgild@localhost ~]$ hdfs dfs -mkdir /user/acadgild/hadoopdata/pig
```

```
18/07/06 16:53:32 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your
```



```

platform... using builtin-java classes where applicable
[acadgild@localhost ~]$ hdfs dfs -mkdir /user/acadgild/hadoopdata/pig/input
18/07/06 16:53:56 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your
platform... using builtin-java classes where applicable
[acadgild@localhost ~]$ hdfs dfs -mkdir /user/acadgild/hadoopdata/pig/output1
18/07/06 16:55:18 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your
platform... using builtin-java classes where applicable
[acadgild@localhost ~]$ hdfs dfs -mkdir /user/acadgild/hadoopdata/pig/output2
18/07/06 16:55:35 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your
platform... using builtin-java classes where applicable
^[A[acadgild@localhost ~]$ hdfs dfs -mkdir /user/acadgild/hadoopdata/pig/output3
18/07/06 16:55:44 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your
platform... using builtin-java classes where applicable
[acadgild@localhost ~]$ hdfs dfs -mkdir /user/acadgild/hadoopdata/pig/output4
18/07/06 16:55:51 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your
platform... using builtin-java classes where applicable
[acadgild@localhost ~]$ hdfs dfs -mkdir /user/acadgild/hadoopdata/pig/output5
18/07/06 16:55:58 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your
platform... using builtin-java classes where applicable
[acadgild@localhost ~]$ hdfs dfs -put /home/acadgild/Desktop/employee_details.txt
/user/acadgild/hadoopdata/pig/input
18/07/06 16:56:38 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your
platform... using builtin-java classes where applicable
[acadgild@localhost ~]$ hdfs dfs -put /home/acadgild/Desktop/employee_expenses.txt
/user/acadgild/hadoopdata/pig/input
18/07/06 16:57:22 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your
platform... using builtin-java classes where applicable
[acadgild@localhost ~]$ hdfs dfs -ls /user/acadgild/hadoopdata/pig/input
18/07/06 16:57:44 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your
platform... using builtin-java classes where applicable
Found 2 items
-rw-r--r-- 1 acadgild supergroup    261 2018-07-06 16:56
/user/acadgild/hadoopdata/pig/input/employee_details.txt
-rw-r--r-- 1 acadgild supergroup     72 2018-07-06 16:57
/user/acadgild/hadoopdata/pig/input/employee_expenses.txt

[acadgild@localhost ~]$ hdfs dfs -cat /user/acadgild/hadoopdata/pig/input/employee_details.txt
18/07/06 17:00:00 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your
platform... using builtin-java classes where applicable
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 ~]$ hdfs dfs -cat /user/acadgild/hadoopdata/pig/input/employee_expenses.txt
18/07/06 17:00:23 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your
platform... using builtin-java classes where applicable
101 200
102 100
110 400
114 200
119 200
105 100
101 100
104 300
102 400

```

**(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)**

**Pig Script : filename: query1.pig**

```

EmpDetail = LOAD '/user/acadgild/hadoopdata/pig/input/employee_details.txt' USING PigStorage(',')
AS(EmpID:int,name:chararray,salary:int,DepartmentID:int);
ExpenseDetail = LOAD '/user/acadgild/hadoopdata/pig/input/employee_expenses.txt' USING
PigStorage(' ') AS (id:int,expense:int);
ratingOrder = ORDER EmpDetail BY DepartmentID asc;
limit5 = limit ratingOrder 5;
SPLIT limit5 into rating1 if DepartmentID == 1, rating2 if DepartmentID == 2;
orderedRating1 = ORDER rating1 by name asc;
Sol1 = UNION rating2,orderedRating1;
SolFinal = FOREACH Sol1 GENERATE EmpID,name;
dump SolFinal;

```

**Terminal Execution:**

```

2018-07-10 23:08:02,449 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter - number of
splits:1
2018-07-10 23:08:02,538 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter -
Submitting tokens for job: job_1531229668339_0065

```

2018-07-10 23:08:02,545 [JobControl] INFO org.apache.hadoop.mapred.YARNRunner - Job jar is not present. Not adding any jar to the list of resources.

2018-07-10 23:08:02,636 [JobControl] INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl - Submitted application application\_1531229668339\_0065

2018-07-10 23:08:02,641 [JobControl] INFO org.apache.hadoop.mapreduce.Job - The url to track the job: [http://localhost:8088/proxy/application\\_1531229668339\\_0065/](http://localhost:8088/proxy/application_1531229668339_0065/)

2018-07-10 23:08:02,649 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - HadoopJobId: job\_1531229668339\_0065

2018-07-10 23:08:02,649 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Processing aliases ratingOrder

2018-07-10 23:08:02,650 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - detailed locations: M: ratingOrder[3,14] C: R:

2018-07-10 23:08:24,815 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 43% complete

2018-07-10 23:08:24,815 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Running jobs are [job\_1531229668339\_0065]

2018-07-10 23:08:32,836 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 50% complete

2018-07-10 23:08:32,836 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Running jobs are [job\_1531229668339\_0065]

2018-07-10 23:08:37,848 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032

2018-07-10 23:08:37,874 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

2018-07-10 23:08:38,140 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032

2018-07-10 23:08:38,156 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

2018-07-10 23:08:38,279 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032

2018-07-10 23:08:38,287 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

2018-07-10 23:08:38,436 [main] INFO org.apache.pig.tools.pigstats.mapreduce.MRScriptState - Pig script settings are added to the job

2018-07-10 23:08:38,436 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - mapred.job.reduce.markreset.buffer.percent is not set, set to default 0.3

2018-07-10 23:08:38,440 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - This job

cannot be converted run in-process

2018-07-10 23:08:38,598 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar file:/home/acadgild/install/pig/pig-0.16.0/pig-0.16.0-core-h2.jar to DistributedCache through /tmp/temp775113582/tmp-1045816564/pig-0.16.0-core-h2.jar

2018-07-10 23:08:38,657 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar file:/home/acadgild/install/pig/pig-0.16.0/lib/automaton-1.11-8.jar to DistributedCache through /tmp/temp775113582/tmp1359779093/automaton-1.11-8.jar

2018-07-10 23:08:38,733 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar file:/home/acadgild/install/pig/pig-0.16.0/lib/antlr-runtime-3.4.jar to DistributedCache through /tmp/temp775113582/tmp1103603524/antlr-runtime-3.4.jar

2018-07-10 23:08:38,790 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar file:/home/acadgild/install/pig/pig-0.16.0/lib/joda-time-2.9.3.jar to DistributedCache through /tmp/temp775113582/tmp562569926/joda-time-2.9.3.jar

2018-07-10 23:08:38,793 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Setting up single store job

2018-07-10 23:08:38,794 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Key [pig.schematuple] is false, will not generate code.

2018-07-10 23:08:38,794 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Starting process to move generated code to distributed cacche

2018-07-10 23:08:38,795 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Setting key [pig.schematuple.classes] with classes to deserialize []

2018-07-10 23:08:38,855 [main] INFO

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

2018-07-10 23:08:38,858 [JobControl] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032

2018-07-10 23:08:38,894 [JobControl] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS

2018-07-10 23:08:38,919 [JobControl] WARN org.apache.hadoop.mapreduce.JobResourceUploader - No job jar file set. User classes may not be found. See Job or Job#setJar(String).

2018-07-10 23:08:38,978 [JobControl] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1

2018-07-10 23:08:38,978 [JobControl] INFO

org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1

2018-07-10 23:08:38,978 [JobControl] INFO

org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths (combined) to process : 1

2018-07-10 23:08:39,074 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter - number of splits:1

2018-07-10 23:08:39,161 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter - Submitting tokens for job: job\_1531229668339\_0066

2018-07-10 23:08:39,169 [JobControl] INFO org.apache.hadoop.mapred.YARNRunner - Job jar is not present. Not adding any jar to the list of resources.

2018-07-10 23:08:39,268 [JobControl] INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl -

Submitted application application\_1531229668339\_0066  
2018-07-10 23:08:39,292 [JobControl] INFO org.apache.hadoop.mapreduce.Job - The url to track the job: http://localhost:8088/proxy/application\_1531229668339\_0066/  
2018-07-10 23:08:39,357 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - HadoopJobId: job\_1531229668339\_0066  
2018-07-10 23:08:39,357 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Processing aliases 1-5, rating1  
2018-07-10 23:08:39,357 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - detailed locations: M: 1-5[-1,-1], rating1[5,18], rating1[-1,-1] C: R:  
2018-07-10 23:08:58,531 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 56% complete  
2018-07-10 23:08:58,531 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Running jobs are [job\_1531229668339\_0066]  
2018-07-10 23:08:59,532 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 62% complete  
2018-07-10 23:08:59,541 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032  
2018-07-10 23:08:59,559 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server  
2018-07-10 23:08:59,835 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032  
2018-07-10 23:08:59,844 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server  
2018-07-10 23:08:59,960 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032  
2018-07-10 23:08:59,966 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server  
2018-07-10 23:09:00,098 [main] INFO org.apache.pig.tools.pigstats.mapreduce.MRScriptState - Pig script settings are added to the job  
2018-07-10 23:09:00,099 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - mapred.job.reduce.markreset.buffer.percent is not set, set to default 0.3  
2018-07-10 23:09:00,100 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Reduce phase detected, estimating # of required reducers.  
2018-07-10 23:09:00,100 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Using reducer estimator:  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.InputSizeReducerEstimator  
2018-07-10 23:09:00,121 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.InputSizeReducerEstimator - BytesPerReducer=1000000000 maxReducers=999 totalInputFileSize=60  
2018-07-10 23:09:00,121 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Setting Parallelism to 1  
2018-07-10 23:09:00,121 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - This job cannot be converted run in-process  
2018-07-10 23:09:00,287 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar file:/home/acadgild/install/pig/pig-0.16.0/pig-0.16.0-core-h2.jar to DistributedCache through /tmp/temp775113582/tmp1373275455/pig-0.16.0-core-h2.jar  
2018-07-10 23:09:00,347 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar file:/home/acadgild/install/pig/pig-0.16.0/lib/automaton-1.11-8.jar to DistributedCache through /tmp/temp775113582/tmp918852207/automaton-1.11-8.jar  
2018-07-10 23:09:00,409 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar file:/home/acadgild/install/pig/pig-0.16.0/lib/antlr-runtime-3.4.jar to DistributedCache through /tmp/temp775113582/tmp-499905557/antlr-runtime-3.4.jar  
2018-07-10 23:09:00,487 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar file:/home/acadgild/install/pig/pig-0.16.0/lib/joda-time-2.9.3.jar to DistributedCache through /tmp/temp775113582/tmp-895941567/joda-time-2.9.3.jar  
2018-07-10 23:09:00,510 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Setting up single store job  
2018-07-10 23:09:00,513 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Key [pig.schematuple] is false, will not generate code.  
2018-07-10 23:09:00,513 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Starting process to move generated code to distributed cacche  
2018-07-10 23:09:00,513 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Setting key [pig.schematuple.classes] with classes to deserialize []  
2018-07-10 23:09:00,597 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 1 map-reduce job(s) waiting for submission.  
2018-07-10 23:09:00,600 [JobControl] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032  
2018-07-10 23:09:00,616 [JobControl] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS  
2018-07-10 23:09:00,645 [JobControl] WARN org.apache.hadoop.mapreduce.JobResourceUploader - No job jar file set. User classes may not be found. See Job or Job#setJar(String).  
2018-07-10 23:09:00,708 [JobControl] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1  
2018-07-10 23:09:00,708 [JobControl] INFO  
org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1  
2018-07-10 23:09:00,709 [JobControl] INFO  
org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths (combined) to process : 1

2018-07-10 23:09:00,826 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter - number of splits:1

2018-07-10 23:09:00,892 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter - Submitting tokens for job: job\_1531229668339\_0067

2018-07-10 23:09:00,901 [JobControl] INFO org.apache.hadoop.mapred.YARNRunner - Job jar is not present. Not adding any jar to the list of resources.

2018-07-10 23:09:00,991 [JobControl] INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl - Submitted application application\_1531229668339\_0067

2018-07-10 23:09:01,001 [JobControl] INFO org.apache.hadoop.mapreduce.Job - The url to track the job: http://localhost:8088/proxy/application\_1531229668339\_0067/

2018-07-10 23:09:01,099 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - HadoopJobId: job\_1531229668339\_0067

2018-07-10 23:09:01,099 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Processing aliases orderedRating1

2018-07-10 23:09:01,099 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - detailed locations: M: orderedRating1[6,17] C: R:

2018-07-10 23:09:23,261 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 68% complete

2018-07-10 23:09:23,262 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Running jobs are [job\_1531229668339\_0067]

2018-07-10 23:09:36,296 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 75% complete

2018-07-10 23:09:36,296 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Running jobs are [job\_1531229668339\_0067]

2018-07-10 23:09:41,331 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032

2018-07-10 23:09:41,380 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

2018-07-10 23:09:41,639 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032

2018-07-10 23:09:41,649 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

2018-07-10 23:09:41,750 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032

2018-07-10 23:09:41,757 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

2018-07-10 23:09:41,839 [main] INFO org.apache.pig.tools.pigstats.mapreduce.MRScriptState - Pig script settings are added to the job

2018-07-10 23:09:41,839 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler -  
mapred.job.reduce.markreset.buffer.percent is not set, set to default 0.3  
2018-07-10 23:09:41,840 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Reduce phase  
detected, estimating # of required reducers.  
2018-07-10 23:09:41,840 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Setting  
Parallelism to 1  
2018-07-10 23:09:41,840 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - This job  
cannot be converted run in-process  
2018-07-10 23:09:41,964 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar  
file:/home/acadgild/install/pig/pig-0.16.0/pig-0.16.0-core-h2.jar to DistributedCache through  
/tmp/temp775113582/tmp1520787764/pig-0.16.0-core-h2.jar  
2018-07-10 23:09:42,066 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar  
file:/home/acadgild/install/pig/pig-0.16.0/lib/automaton-1.11-8.jar to DistributedCache through  
/tmp/temp775113582/tmp-568654101/automaton-1.11-8.jar  
2018-07-10 23:09:42,112 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar  
file:/home/acadgild/install/pig/pig-0.16.0/lib/antlr-runtime-3.4.jar to DistributedCache through  
/tmp/temp775113582/tmp1787619297/antlr-runtime-3.4.jar  
2018-07-10 23:09:42,176 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar  
file:/home/acadgild/install/pig/pig-0.16.0/lib/joda-time-2.9.3.jar to DistributedCache through  
/tmp/temp775113582/tmp-1020137455/joda-time-2.9.3.jar  
2018-07-10 23:09:42,183 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Setting up  
single store job  
2018-07-10 23:09:42,185 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Key  
[pig.schematuple] is false, will not generate code.  
2018-07-10 23:09:42,185 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Starting process  
to move generated code to distributed cacche  
2018-07-10 23:09:42,185 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Setting key  
[pig.schematuple.classes] with classes to deserialize []  
2018-07-10 23:09:42,223 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 1  
map-reduce job(s) waiting for submission.  
2018-07-10 23:09:42,226 [JobControl] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to  
ResourceManager at localhost/127.0.0.1:8032  
2018-07-10 23:09:42,247 [JobControl] INFO org.apache.hadoop.conf.Configuration.deprecation -  
fs.default.name is deprecated. Instead, use fs.defaultFS  
2018-07-10 23:09:42,274 [JobControl] WARN org.apache.hadoop.mapreduce.JobResourceUploader -  
No job jar file set. User classes may not be found. See Job or Job#setJar(String).  
2018-07-10 23:09:42,345 [JobControl] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat  
- Total input paths to process : 1  
2018-07-10 23:09:42,345 [JobControl] INFO  
org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1



2018-07-10 23:09:42,345 [JobControl] INFO  
org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths (combined) to  
process : 1

2018-07-10 23:09:42,461 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter - number of  
splits:1

2018-07-10 23:09:42,561 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter -  
Submitting tokens for job: job\_1531229668339\_0068

2018-07-10 23:09:42,573 [JobControl] INFO org.apache.hadoop.mapred.YARNRunner - Job jar is not  
present. Not adding any jar to the list of resources.

2018-07-10 23:09:42,675 [JobControl] INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl -  
Submitted application application\_1531229668339\_0068

2018-07-10 23:09:42,684 [JobControl] INFO org.apache.hadoop.mapreduce.Job - The url to track the  
job: [http://localhost:8088/proxy/application\\_1531229668339\\_0068/](http://localhost:8088/proxy/application_1531229668339_0068/)

2018-07-10 23:09:42,725 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher -  
HadoopJobId: job\_1531229668339\_0068

2018-07-10 23:09:42,725 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Processing  
aliases orderedRating1

2018-07-10 23:09:42,726 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - detailed  
locations: M: orderedRating1[6,17] C: R: orderedRating1[-1,-1]

2018-07-10 23:10:05,006 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 81%  
complete

2018-07-10 23:10:05,006 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Running  
jobs are [job\_1531229668339\_0068]

2018-07-10 23:10:15,028 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 87%  
complete

2018-07-10 23:10:15,028 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Running  
jobs are [job\_1531229668339\_0068]

2018-07-10 23:10:18,036 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to  
ResourceManager at localhost/127.0.0.1:8032

2018-07-10 23:10:18,050 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate -  
Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history  
server

2018-07-10 23:10:18,301 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to  
ResourceManager at localhost/127.0.0.1:8032

2018-07-10 23:10:18,321 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate -  
Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history  
server

2018-07-10 23:10:18,473 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to  
ResourceManager at localhost/127.0.0.1:8032

2018-07-10 23:10:18,488 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate -  
Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history  
server

2018-07-10 23:10:18,586 [main] INFO org.apache.pig.tools.pigstats.mapreduce.MRScriptState - Pig script settings are added to the job

2018-07-10 23:10:18,587 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - mapred.job.reduce.markreset.buffer.percent is not set, set to default 0.3

2018-07-10 23:10:18,590 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - This job cannot be converted run in-process

2018-07-10 23:10:18,894 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar file:/home/acadgild/install/pig/pig-0.16.0/pig-0.16.0-core-h2.jar to DistributedCache through /tmp/temp775113582/tmp-867225340/pig-0.16.0-core-h2.jar

2018-07-10 23:10:19,027 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar file:/home/acadgild/install/pig/pig-0.16.0/lib/automaton-1.11-8.jar to DistributedCache through /tmp/temp775113582/tmp2008461921/automaton-1.11-8.jar

2018-07-10 23:10:19,130 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar file:/home/acadgild/install/pig/pig-0.16.0/lib/antlr-runtime-3.4.jar to DistributedCache through /tmp/temp775113582/tmp-387182359/antlr-runtime-3.4.jar

2018-07-10 23:10:19,195 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar file:/home/acadgild/install/pig/pig-0.16.0/lib/joda-time-2.9.3.jar to DistributedCache through /tmp/temp775113582/tmp-1065646931/joda-time-2.9.3.jar

2018-07-10 23:10:19,198 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Setting up single store job

2018-07-10 23:10:19,199 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Key [pig.schematuple] is false, will not generate code.

2018-07-10 23:10:19,199 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Starting process to move generated code to distributed cacche

2018-07-10 23:10:19,199 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Setting key [pig.schematuple.classes] with classes to deserialize []

2018-07-10 23:10:19,279 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 1 map-reduce job(s) waiting for submission.

2018-07-10 23:10:19,281 [JobControl] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032

2018-07-10 23:10:19,300 [JobControl] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS

2018-07-10 23:10:19,324 [JobControl] WARN org.apache.hadoop.mapreduce.JobResourceUploader - No job jar file set. User classes may not be found. See Job or Job#setJar(String).

2018-07-10 23:10:19,374 [JobControl] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1

2018-07-10 23:10:19,374 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1

2018-07-10 23:10:19,374 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths (combined) to process : 1

2018-07-10 23:10:19,392 [JobControl] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat  
- Total input paths to process : 1  
2018-07-10 23:10:19,392 [JobControl] INFO  
org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1  
2018-07-10 23:10:19,392 [JobControl] INFO  
org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths (combined) to  
process : 1  
2018-07-10 23:10:19,475 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter - number of  
splits:2  
2018-07-10 23:10:19,525 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter -  
Submitting tokens for job: job\_1531229668339\_0069  
2018-07-10 23:10:19,529 [JobControl] INFO org.apache.hadoop.mapred.YARNRunner - Job jar is not  
present. Not adding any jar to the list of resources.  
2018-07-10 23:10:19,596 [JobControl] INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl -  
Submitted application application\_1531229668339\_0069  
2018-07-10 23:10:19,602 [JobControl] INFO org.apache.hadoop.mapreduce.Job - The url to track the  
job: http://localhost:8088/proxy/application\_1531229668339\_0069/  
2018-07-10 23:10:19,780 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher -  
HadoopJobId: job\_1531229668339\_0069  
2018-07-10 23:10:19,780 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Processing  
aliases 1-5,Sol1,SolFinal,rating2  
2018-07-10 23:10:19,780 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - detailed  
locations: M: 1-5[-1,-1],rating2[5,48],rating2[-1,-1],Sol1[7,7],SolFinal[8,11] C: R:  
2018-07-10 23:10:48,840 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 93%  
complete  
2018-07-10 23:10:48,840 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Running  
jobs are [job\_1531229668339\_0069]  
2018-07-10 23:10:54,860 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to  
ResourceManager at localhost/127.0.0.1:8032  
2018-07-10 23:10:54,870 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate -  
Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history  
server  
2018-07-10 23:10:55,024 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to  
ResourceManager at localhost/127.0.0.1:8032  
2018-07-10 23:10:55,042 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate -  
Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history  
server  
2018-07-10 23:10:55,168 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to  
ResourceManager at localhost/127.0.0.1:8032  
2018-07-10 23:10:55,185 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate -  
Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history  
server  
2018-07-10 23:10:55,342 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 100%

complete

2018-07-10 23:10:55,840 [main] INFO org.apache.pig.tools.pigstats.mapreduce.SimplePigStats -  
Script Statistics:

HadoopVersion	PigVersion	UserId	StartedAt	FinishedAt	Features
2.6.5	0.16.0	acadgild	2018-07-10 23:06:29	2018-07-10 23:10:55	ORDER_BY,LIMIT,UNION

Success!

Job Stats (time in seconds):

JobId	Maps	Reduces	MaxMapTime	MinMapTime	AvgMapTime	MedianMapTime	MaxReduceTime	MinReduceTime	AvgReduceTime	MedianReductime	Alias	Feature
Outputs												
job_1531229668339_0062	1	0	6	6	6	6	0	0	0	0	0	0
EmpDetail	MAP_ONLY											
job_1531229668339_0063	1	1	6	6	6	6	6	6	6	6	6	6
ratingOrder	SAMPLER											
job_1531229668339_0064	1	1	6	6	6	6	6	6	6	6	6	6
ratingOrder	ORDER_BY,COMBINER											
job_1531229668339_0065	1	1	5	5	5	5	6	6	6	6	6	6
ratingOrder												
job_1531229668339_0066	1	0	6	6	6	6	0	0	0	0	0	0
1-5,rating1	MAP_ONLY											
job_1531229668339_0067	1	1	6	6	6	6	9	9	9	9	9	9
orderedRating1	SAMPLER											
job_1531229668339_0068	1	1	6	6	6	6	7	7	7	7	7	7
orderedRating1	ORDER_BY											
job_1531229668339_0069	2	0	14	14	14	14	0	0	0	0	0	0
1-5,Sol1,SolFinal,rating2	MAP_ONLY											
hdfs://localhost:8020/tmp/temp775113582/tmp-341850608												

,

Input(s):

Successfully read 14 records (665 bytes) from:  
"/user/acadgild/hadoopdata/pig/input/employee\_details.txt"

Output(s):

Successfully stored 5 records (76 bytes) in:  
"hdfs://localhost:8020/tmp/temp775113582/tmp-341850608"

Counters:

Total records written : 5

Total bytes written : 76

Spillable Memory Manager spill count : 0

Total bags proactively spilled: 0

Total records proactively spilled: 0

Job DAG:

job\_1531229668339\_0062 -> job\_1531229668339\_0063,  
job\_1531229668339\_0063 -> job\_1531229668339\_0064,

job\_1531229668339\_0064 -> job\_1531229668339\_0065,  
job\_1531229668339\_0065 -> job\_1531229668339\_0066,  
job\_1531229668339\_0066 -> job\_1531229668339\_0067,  
job\_1531229668339\_0067 -> job\_1531229668339\_0068,  
job\_1531229668339\_0068 -> job\_1531229668339\_0069,  
job\_1531229668339\_0069

2018-07-10 23:10:55,843 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032

2018-07-10 23:10:55,847 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

2018-07-10 23:10:55,955 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032

2018-07-10 23:10:55,970 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

2018-07-10 23:10:56,052 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032

2018-07-10 23:10:56,058 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

2018-07-10 23:10:56,156 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032

2018-07-10 23:10:56,164 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

2018-07-10 23:10:56,281 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032

2018-07-10 23:10:56,297 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

2018-07-10 23:10:56,379 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032

2018-07-10 23:10:56,389 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

2018-07-10 23:10:56,492 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032

2018-07-10 23:10:56,500 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

2018-07-10 23:10:56,616 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032

2018-07-10 23:10:56,621 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

2018-07-10 23:10:56,700 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to

ResourceManager at localhost/127.0.0.1:8032

2018-07-10 23:10:56,704 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

2018-07-10 23:10:56,907 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032

2018-07-10 23:10:56,912 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

2018-07-10 23:10:57,046 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032

2018-07-10 23:10:57,063 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

2018-07-10 23:10:57,165 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032

2018-07-10 23:10:57,172 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

2018-07-10 23:10:57,262 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032

2018-07-10 23:10:57,268 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

2018-07-10 23:10:57,392 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032

2018-07-10 23:10:57,397 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

2018-07-10 23:10:57,517 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032

2018-07-10 23:10:57,525 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

2018-07-10 23:10:57,609 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032

2018-07-10 23:10:57,618 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

2018-07-10 23:10:57,737 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032

2018-07-10 23:10:57,744 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

2018-07-10 23:10:57,826 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032

2018-07-10 23:10:57,835 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

2018-07-10 23:10:57,932 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032

2018-07-10 23:10:57,937 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

2018-07-10 23:10:58,044 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032

2018-07-10 23:10:58,049 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

2018-07-10 23:10:58,175 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032

2018-07-10 23:10:58,182 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

2018-07-10 23:10:58,288 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032

2018-07-10 23:10:58,298 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

2018-07-10 23:10:58,486 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032

2018-07-10 23:10:58,504 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

2018-07-10 23:10:58,582 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032

2018-07-10 23:10:58,587 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

2018-07-10 23:10:58,668 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!

2018-07-10 23:10:58,679 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS

2018-07-10 23:10:58,684 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.

2018-07-10 23:10:58,695 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 2

2018-07-10 23:10:58,695 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 2  
(106,Aamir)  
(101,Amitabh)  
(113,Jubeen)  
(111,Tushar)  
(114,Madhuri)

2018-07-10 23:10:59,207 [main] INFO org.apache.pig.Main - Pig script completed in 4 minutes, 36 seconds and 366 milliseconds (276366 ms)

[acadgild@localhost ~]\$

**(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)**

**Pig Script : filename: query2.pig**

```
EmpDetail = LOAD '/user/acadgild/hadoopdata/pig/input/employee_details.txt' USING PigStorage(',')
AS(EmpID:int,name:chararray,salary:int,DepartmentID:int);
ExpenseDetail = LOAD '/user/acadgild/hadoopdata/pig/input/employee_expenses.txt' USING
PigStorage(' ') AS (id:int,expense:int);
odd = FILTER EmpDetail BY (EmpID%2 != 0);
O = ORDER odd BY salary desc;
limit3 = limit O 3;
sol2 = FOREACH limit3 GENERATE EmpID,name;
dump sol2;
```

**Terminal Execution:**

```
[acadgild@localhost ~]$ pig /home/acadgild/Desktop/query2.pig
[acadgild@localhost ~]$ pig /home/acadgild/Desktop/query2.pig
18/07/10 23:20:58 INFO pig.ExecTypeProvider: Trying ExecType : LOCAL
18/07/10 23:20:58 INFO pig.ExecTypeProvider: Trying ExecType : MAPREDUCE
18/07/10 23:20:58 INFO pig.ExecTypeProvider: Picked MAPREDUCE as the ExecType
2018-07-10 23:20:58,321 [main] INFO org.apache.pig.Main - Apache Pig version 0.16.0 (r1746530)
compiled Jun 01 2016, 23:10:49
2018-07-10 23:20:58,321 [main] INFO org.apache.pig.Main - Logging error messages to:
/home/acadgild/pig_1531245058318.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-1.4.4/lib/slf4j-log4j12-1.7.10.jar!/org/slf4j/impl/StaticLogg
erBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
2018-07-10 23:20:59,251 [main] WARN org.apache.hadoop.util.NativeCodeLoader - Unable to load
native-hadoop library for your platform... using builtin-java classes where applicable
2018-07-10 23:20:59,645 [main] INFO org.apache.pig.impl.util.Utils - Default bootup file
/home/acadgild/.pigbootup not found
```



2018-07-10 23:20:59,878 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address

2018-07-10 23:20:59,878 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS

2018-07-10 23:20:59,881 [main] INFO org.apache.pig.backend.hadoop.executionengine.HExecutionEngine - Connecting to hadoop file system at: hdfs://localhost:8020

2018-07-10 23:21:00,978 [main] INFO org.apache.pig.PigServer - Pig Script ID for the session: PIG-query2.pig-7ee1b67f-1c65-440c-a31d-4fc8906c3995

2018-07-10 23:21:00,978 [main] WARN org.apache.pig.PigServer - ATS is disabled since yarn.timeline-service.enabled set to false

2018-07-10 23:21:02,714 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS

2018-07-10 23:21:02,927 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS

2018-07-10 23:21:03,388 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: ORDER\_BY,FILTER,LIMIT

2018-07-10 23:21:03,494 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS

2018-07-10 23:21:03,514 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.

2018-07-10 23:21:03,651 [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]}

2018-07-10 23:21:03,739 [main] INFO org.apache.pig.newplan.logical.rules.ColumnPruneVisitor - Columns pruned for EmpDetail: \$3

2018-07-10 23:21:03,888 [main] INFO org.apache.pig.impl.util.SpillableMemoryManager - Selected heap (Tenured Gen) of size 699072512 to monitor. collectionUsageThreshold = 489350752, usageThreshold = 489350752

2018-07-10 23:21:04,072 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File concatenation threshold: 100 optimistic? false

2018-07-10 23:21:04,211 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.SecondaryKeyOptimizerMR - Using Secondary Key Optimization for MapReduce node scope-33

2018-07-10 23:21:04,278 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size before optimization: 4

2018-07-10 23:21:04,278 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size after optimization: 4

2018-07-10 23:21:04,379 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS

2018-07-10 23:21:04,485 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032

2018-07-10 23:21:05,085 [main] INFO org.apache.pig.tools.pigstats.mapreduce.MRScriptState - Pig

script settings are added to the job

2018-07-10 23:21:05,106 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.reduce.markreset.buffer.percent is deprecated. Instead, use mapreduce.reduce.markreset.buffer.percent

2018-07-10 23:21:05,106 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - mapred.job.reduce.markreset.buffer.percent is not set, set to default 0.3

2018-07-10 23:21:05,106 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.output.compress is deprecated. Instead, use mapreduce.output.fileoutputformat.compress

2018-07-10 23:21:05,119 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - This job cannot be converted run in-process

2018-07-10 23:21:05,180 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.submit.replication is deprecated. Instead, use mapreduce.client.submit.file.replication

2018-07-10 23:21:05,893 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar file:/home/acadgild/install/pig/pig-0.16.0/pig-0.16.0-core-h2.jar to DistributedCache through /tmp/temp856022726/tmp-602511734/pig-0.16.0-core-h2.jar

2018-07-10 23:21:05,994 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar file:/home/acadgild/install/pig/pig-0.16.0/lib/automaton-1.11-8.jar to DistributedCache through /tmp/temp856022726/tmp-388919429/automaton-1.11-8.jar

2018-07-10 23:21:06,067 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar file:/home/acadgild/install/pig/pig-0.16.0/lib/antlr-runtime-3.4.jar to DistributedCache through /tmp/temp856022726/tmp501549568/antlr-runtime-3.4.jar

2018-07-10 23:21:06,180 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar file:/home/acadgild/install/pig/pig-0.16.0/lib/joda-time-2.9.3.jar to DistributedCache through /tmp/temp856022726/tmp-688000832/joda-time-2.9.3.jar

2018-07-10 23:21:06,215 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Setting up single store job

2018-07-10 23:21:06,266 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Key [pig.schematuple] is false, will not generate code.

2018-07-10 23:21:06,266 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Starting process to move generated code to distributed cacche

2018-07-10 23:21:06,267 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Setting key [pig.schematuple.classes] with classes to deserialize []

2018-07-10 23:21:06,404 [main] INFO

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

2018-07-10 23:21:06,405 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker.http.address is deprecated. Instead, use mapreduce.jobtracker.http.address

2018-07-10 23:21:06,412 [JobControl] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032

2018-07-10 23:21:06,470 [JobControl] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.task.id is deprecated. Instead, use mapreduce.task.attempt.id

2018-07-10 23:21:06,474 [JobControl] INFO org.apache.hadoop.conf.Configuration.deprecation -

fs.default.name is deprecated. Instead, use fs.defaultFS  
2018-07-10 23:21:07,120 [JobControl] WARN org.apache.hadoop.mapreduce.JobResourceUploader - No job jar file set. User classes may not be found. See Job or Job#setJar(String).  
2018-07-10 23:21:07,211 [JobControl] INFO org.apache.pig.builtin.PigStorage - Using PigTextInputFormat  
2018-07-10 23:21:07,229 [JobControl] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1  
2018-07-10 23:21:07,229 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1  
2018-07-10 23:21:07,350 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths (combined) to process : 1  
2018-07-10 23:21:07,527 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter - number of splits:1  
2018-07-10 23:21:07,850 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter - Submitting tokens for job: job\_1531229668339\_0070  
2018-07-10 23:21:08,272 [JobControl] INFO org.apache.hadoop.mapred.YARNRunner - Job jar is not present. Not adding any jar to the list of resources.  
2018-07-10 23:21:08,516 [JobControl] INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl - Submitted application application\_1531229668339\_0070  
2018-07-10 23:21:08,635 [JobControl] INFO org.apache.hadoop.mapreduce.Job - The url to track the job: http://localhost:8088/proxy/application\_1531229668339\_0070/  
2018-07-10 23:21:08,635 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - HadoopJobId: job\_1531229668339\_0070  
2018-07-10 23:21:08,635 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Processing aliases EmpDetail,odd  
2018-07-10 23:21:08,635 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - detailed locations: M: EmpDetail[1,12],EmpDetail[-1,-1],odd[3,6] C: R:  
2018-07-10 23:21:08,662 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 0% complete  
2018-07-10 23:21:08,662 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Running jobs are [job\_1531229668339\_0070]  
2018-07-10 23:21:33,022 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 12% complete  
2018-07-10 23:21:33,022 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Running jobs are [job\_1531229668339\_0070]  
2018-07-10 23:21:39,035 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 25% complete  
2018-07-10 23:21:39,041 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032  
2018-07-10 23:21:39,067 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate -

Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

2018-07-10 23:21:39,482 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032

2018-07-10 23:21:39,490 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

2018-07-10 23:21:39,698 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.reduce.tasks is deprecated. Instead, use mapreduce.job.reduces

2018-07-10 23:21:39,699 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032

2018-07-10 23:21:39,709 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

2018-07-10 23:21:39,847 [main] INFO org.apache.pig.tools.pigstats.mapreduce.MRScriptState - Pig script settings are added to the job

2018-07-10 23:21:39,848 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - mapred.job.reduce.markreset.buffer.percent is not set, set to default 0.3

2018-07-10 23:21:39,850 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Reduce phase detected, estimating # of required reducers.

2018-07-10 23:21:39,852 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Using reducer estimator:

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.InputSizeReducerEstimator

2018-07-10 23:21:39,899 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.InputSizeReducerEstimator - BytesPerReducer=1000000000 maxReducers=999 totalInputFileSize=127

2018-07-10 23:21:39,903 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Setting Parallelism to 1

2018-07-10 23:21:39,903 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - This job cannot be converted run in-process

2018-07-10 23:21:40,050 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar file:/home/acadgild/install/pig/pig-0.16.0/pig-0.16.0-core-h2.jar to DistributedCache through /tmp/temp856022726/tmp-211910625/pig-0.16.0-core-h2.jar

2018-07-10 23:21:40,121 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar file:/home/acadgild/install/pig/pig-0.16.0/lib/automaton-1.11-8.jar to DistributedCache through /tmp/temp856022726/tmp-978427596/automaton-1.11-8.jar

2018-07-10 23:21:40,172 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar file:/home/acadgild/install/pig/pig-0.16.0/lib/antlr-runtime-3.4.jar to DistributedCache through /tmp/temp856022726/tmp-1830235691/antlr-runtime-3.4.jar

2018-07-10 23:21:40,273 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar

file:/home/acadgild/install/pig/pig-0.16.0/lib/joda-time-2.9.3.jar to DistributedCache through  
/tmp/temp856022726/tmp1892965328/joda-time-2.9.3.jar  
2018-07-10 23:21:40,279 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Setting up  
single store job  
2018-07-10 23:21:40,289 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Key  
[pig.schematuple] is false, will not generate code.  
2018-07-10 23:21:40,289 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Starting process  
to move generated code to distributed cacche  
2018-07-10 23:21:40,289 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Setting key  
[pig.schematuple.classes] with classes to deserialize []  
2018-07-10 23:21:40,480 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 1  
map-reduce job(s) waiting for submission.  
2018-07-10 23:21:40,495 [JobControl] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to  
ResourceManager at localhost/127.0.0.1:8032  
2018-07-10 23:21:40,525 [JobControl] INFO org.apache.hadoop.conf.Configuration.deprecation -  
fs.default.name is deprecated. Instead, use fs.defaultFS  
2018-07-10 23:21:40,570 [JobControl] WARN org.apache.hadoop.mapreduce.JobResourceUploader -  
No job jar file set. User classes may not be found. See Job or Job#setJar(String).  
2018-07-10 23:21:40,632 [JobControl] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat  
- Total input paths to process : 1  
2018-07-10 23:21:40,632 [JobControl] INFO  
org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1  
2018-07-10 23:21:40,633 [JobControl] INFO  
org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths (combined) to  
process : 1  
2018-07-10 23:21:40,746 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter - number of  
splits:1  
2018-07-10 23:21:40,870 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter -  
Submitting tokens for job: job\_1531229668339\_0071  
2018-07-10 23:21:40,879 [JobControl] INFO org.apache.hadoop.mapred.YARNRunner - Job jar is not  
present. Not adding any jar to the list of resources.  
2018-07-10 23:21:40,996 [JobControl] INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl -  
Submitted application application\_1531229668339\_0071  
2018-07-10 23:21:41,010 [JobControl] INFO org.apache.hadoop.mapreduce.Job - The url to track the  
job: http://localhost:8088/proxy/application\_1531229668339\_0071/  
2018-07-10 23:21:41,013 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher -  
HadoopJobId: job\_1531229668339\_0071  
2018-07-10 23:21:41,013 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Processing  
aliases O  
2018-07-10 23:21:41,013 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - detailed  
locations: M: O[4,4] C: R:  
2018-07-10 23:21:40,632 [JobControl] INFO  
org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1  
2018-07-10 23:21:40,633 [JobControl] INFO

org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths (combined) to process : 1  
2018-07-10 23:21:40,746 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter - number of splits:1  
2018-07-10 23:21:40,870 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter - Submitting tokens for job: job\_1531229668339\_0071  
2018-07-10 23:21:40,879 [JobControl] INFO org.apache.hadoop.mapred.YARNRunner - Job jar is not present. Not adding any jar to the list of resources.  
2018-07-10 23:21:40,996 [JobControl] INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl - Submitted application application\_1531229668339\_0071  
2018-07-10 23:21:41,010 [JobControl] INFO org.apache.hadoop.mapreduce.Job - The url to track the job: http://localhost:8088/proxy/application\_1531229668339\_0071/  
2018-07-10 23:21:41,013 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - HadoopJobId: job\_1531229668339\_0071  
2018-07-10 23:21:41,013 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Processing aliases O  
2018-07-10 23:21:41,013 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - detailed locations: M: O[4,4] C: R:  
2018-07-10 23:22:06,321 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 37% complete  
2018-07-10 23:22:06,322 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Running jobs are [job\_1531229668339\_0071]  
2018-07-10 23:22:18,353 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 50% complete  
2018-07-10 23:22:18,353 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Running jobs are [job\_1531229668339\_0071]  
2018-07-10 23:22:21,365 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032  
2018-07-10 23:22:21,391 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server  
2018-07-10 23:22:21,618 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032  
2018-07-10 23:22:21,625 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server  
2018-07-10 23:22:21,748 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032  
2018-07-10 23:22:21,768 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server  
2018-07-10 23:22:21,886 [main] INFO org.apache.pig.tools.pigstats.mapreduce.MRScriptState - Pig

script settings are added to the job  
2018-07-10 23:22:21,886 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - mapred.job.reduce.markreset.buffer.percent is not set, set to default 0.3  
2018-07-10 23:22:21,887 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Reduce phase detected, estimating # of required reducers.  
2018-07-10 23:22:21,893 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Setting Parallelism to 1  
2018-07-10 23:22:21,893 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - This job cannot be converted run in-process  
2018-07-10 23:22:22,205 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar file:/home/acadgild/install/pig/pig-0.16.0/pig-0.16.0-core-h2.jar to DistributedCache through /tmp/temp856022726/tmp1942299511/pig-0.16.0-core-h2.jar  
2018-07-10 23:22:22,309 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar file:/home/acadgild/install/pig/pig-0.16.0/lib/automaton-1.11-8.jar to DistributedCache through /tmp/temp856022726/tmp-1317825112/automaton-1.11-8.jar  
2018-07-10 23:22:22,428 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar file:/home/acadgild/install/pig/pig-0.16.0/lib/antlr-runtime-3.4.jar to DistributedCache through /tmp/temp856022726/tmp1103653153/antlr-runtime-3.4.jar  
2018-07-10 23:22:22,516 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar file:/home/acadgild/install/pig/pig-0.16.0/lib/joda-time-2.9.3.jar to DistributedCache through /tmp/temp856022726/tmp-588516278/joda-time-2.9.3.jar  
2018-07-10 23:22:22,566 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Setting up single store job  
2018-07-10 23:22:22,569 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Key [pig.schematuple] is false, will not generate code.  
2018-07-10 23:22:22,569 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Starting process to move generated code to distributed cacche  
2018-07-10 23:22:22,569 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Setting key [pig.schematuple.classes] with classes to deserialize []  
2018-07-10 23:22:22,731 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 1 map-reduce job(s) waiting for submission.  
2018-07-10 23:22:22,734 [JobControl] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032  
2018-07-10 23:22:22,757 [JobControl] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS  
2018-07-10 23:22:22,774 [JobControl] WARN org.apache.hadoop.mapreduce.JobResourceUploader - No job jar file set. User classes may not be found. See Job or Job#setJar(String).  
2018-07-10 23:22:22,819 [JobControl] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1

2018-07-10 23:22:22,819 [JobControl] INFO  
org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1

2018-07-10 23:22:22,819 [JobControl] INFO  
org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths (combined) to process : 1

2018-07-10 23:22:22,905 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter - number of splits:1

2018-07-10 23:22:23,003 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter - Submitting tokens for job: job\_1531229668339\_0072

2018-07-10 23:22:23,010 [JobControl] INFO org.apache.hadoop.mapred.YARNRunner - Job jar is not present. Not adding any jar to the list of resources.

2018-07-10 23:22:23,094 [JobControl] INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl - Submitted application application\_1531229668339\_0072

2018-07-10 23:22:23,105 [JobControl] INFO org.apache.hadoop.mapreduce.Job - The url to track the job: [http://localhost:8088/proxy/application\\_1531229668339\\_0072/](http://localhost:8088/proxy/application_1531229668339_0072/)

2018-07-10 23:22:23,232 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - HadoopJobId: job\_1531229668339\_0072

2018-07-10 23:22:23,233 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Processing aliases O

2018-07-10 23:22:23,233 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - detailed locations: M: O[4,4] C: R:

2018-07-10 23:22:43,532 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 62% complete

2018-07-10 23:22:43,532 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Running jobs are [job\_1531229668339\_0072]

2018-07-10 23:22:57,573 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 75% complete

2018-07-10 23:22:57,573 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Running jobs are [job\_1531229668339\_0072]

2018-07-10 23:22:58,589 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032

2018-07-10 23:22:58,619 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

2018-07-10 23:22:58,802 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032

2018-07-10 23:22:58,810 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

2018-07-10 23:22:58,907 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032

2018-07-10 23:22:58,913 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate -



Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

2018-07-10 23:22:58,986 [main] INFO org.apache.pig.tools.pigstats.mapreduce.MRScriptState - Pig script settings are added to the job

2018-07-10 23:22:58,987 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - mapred.job.reduce.markreset.buffer.percent is not set, set to default 0.3

2018-07-10 23:22:58,988 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Reduce phase detected, estimating # of required reducers.

2018-07-10 23:22:58,988 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Setting Parallelism to 1

2018-07-10 23:22:58,988 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - This job cannot be converted run in-process

2018-07-10 23:22:59,147 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar file:/home/acadgild/install/pig/pig-0.16.0/pig-0.16.0-core-h2.jar to DistributedCache through /tmp/temp856022726/tmp680854566/pig-0.16.0-core-h2.jar

2018-07-10 23:22:59,236 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar file:/home/acadgild/install/pig/pig-0.16.0/lib/automaton-1.11-8.jar to DistributedCache through /tmp/temp856022726/tmp1290510360/automaton-1.11-8.jar

2018-07-10 23:22:59,293 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar file:/home/acadgild/install/pig/pig-0.16.0/lib/antlr-runtime-3.4.jar to DistributedCache through /tmp/temp856022726/tmp934202400/antlr-runtime-3.4.jar

2018-07-10 23:22:59,383 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar file:/home/acadgild/install/pig/pig-0.16.0/lib/joda-time-2.9.3.jar to DistributedCache through /tmp/temp856022726/tmp332134010/joda-time-2.9.3.jar

2018-07-10 23:22:59,385 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Setting up single store job

2018-07-10 23:22:59,386 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Key [pig.schematuple] is false, will not generate code.

2018-07-10 23:22:59,386 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Starting process to move generated code to distributed cacche

2018-07-10 23:22:59,386 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Setting key [pig.schematuple.classes] with classes to deserialize []

2018-07-10 23:22:59,424 [main] INFO

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

2018-07-10 23:22:59,427 [JobControl] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032

2018-07-10 23:22:59,439 [JobControl] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS

2018-07-10 23:22:59,464 [JobControl] WARN org.apache.hadoop.mapreduce.JobResourceUploader -

No job jar file set. User classes may not be found. See Job or Job#setJar(String).

2018-07-10 23:22:59,510 [JobControl] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1

2018-07-10 23:22:59,510 [JobControl] INFO

org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1

2018-07-10 23:22:59,510 [JobControl] INFO

org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths (combined) to process : 1

2018-07-10 23:22:59,624 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter - number of splits:1

2018-07-10 23:22:59,712 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter -

Submitting tokens for job: job\_1531229668339\_0073

2018-07-10 23:22:59,718 [JobControl] INFO org.apache.hadoop.mapred.YARNRunner - Job jar is not present. Not adding any jar to the list of resources.

2018-07-10 23:22:59,805 [JobControl] INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl - Submitted application application\_1531229668339\_0073

2018-07-10 23:22:59,811 [JobControl] INFO org.apache.hadoop.mapreduce.Job - The url to track the job: http://localhost:8088/proxy/application\_1531229668339\_0073/

2018-07-10 23:22:59,926 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - HadoopJobId: job\_1531229668339\_0073

2018-07-10 23:22:59,926 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Processing aliases O,sol2

2018-07-10 23:22:59,926 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - detailed locations: M: O[4,4] C: R: O[-1,-1],sol2[6,7]

2018-07-10 23:23:36,538 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032

2018-07-10 23:23:36,543 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

2018-07-10 23:23:36,662 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032

2018-07-10 23:23:36,668 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

2018-07-10 23:23:36,765 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032

2018-07-10 23:23:36,774 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

2018-07-10 23:23:36,867 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032

2018-07-10 23:23:36,873 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

2018-07-10 23:23:36,973 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032

2018-07-10 23:23:36,983 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

2018-07-10 23:23:37,078 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032

2018-07-10 23:23:37,083 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

2018-07-10 23:23:37,183 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!

2018-07-10 23:23:37,189 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS

2018-07-10 23:23:37,192 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.

2018-07-10 23:23:37,219 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1

2018-07-10 23:23:37,219 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1

**(101,Amitabh)**

**(107,Salman)**

**(103,Akshay)**

2018-07-10 23:23:37,514 [main] INFO org.apache.pig.Main - Pig script completed in 2 minutes, 39 seconds and 658 milliseconds (159658 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)**

**Pig Script : filename: query3.pig**

```
EmpDetail = LOAD '/user/acadgild/hadoopdata/pig/input/employee_details.txt' USING PigStorage(',')
AS(EmpID:int,name:chararray,salary:int,DepartmentID:int);
ExpenseDetail = LOAD '/user/acadgild/hadoopdata/pig/input/employee_expenses.txt' USING
PigStorage(' ') AS (id:int,expense:int);
commonExpense = JOIN EmpDetail by EmpID RIGHT, ExpenseDetail by id;
expensesOrder = ORDER commonExpense BY expense desc;
limit1 = LIMIT expensesOrder 1;
expenseMaxEmp = FOREACH limit1 GENERATE EmpID,name;
dump expenseMaxEmp;
```

## **Terminal Execution:**

```
[acadgild@localhost ~]$ pig /home/acadgild/Desktop/query3.pig
18/07/10 23:25:07 INFO pig.ExecTypeProvider: Trying ExecType : LOCAL
18/07/10 23:25:07 INFO pig.ExecTypeProvider: Trying ExecType : MAPREDUCE
18/07/10 23:25:07 INFO pig.ExecTypeProvider: Picked MAPREDUCE as the ExecType
2018-07-10 23:25:07,905 [main] INFO  org.apache.pig.Main - Apache Pig version 0.16.0 (r1746530)
compiled Jun 01 2016, 23:10:49
2018-07-10 23:25:07,905 [main] INFO  org.apache.pig.Main - Logging error messages to:
/home/acadgild/pig_1531245307902.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-1.4.4/lib/slf4j-log4j12-1.7.10.jar!/org/slf4j/impl/StaticLogg
erBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
2018-07-10 23:25:09,016 [main] WARN  org.apache.hadoop.util.NativeCodeLoader - Unable to load
native-hadoop library for your platform... using builtin-java classes where applicable
2018-07-10 23:25:09,503 [main] INFO  org.apache.pig.impl.util.Utils - Default bootup file
/home/acadgild/.pigbootup not found
2018-07-10 23:25:09,761 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation -
mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2018-07-10 23:25:09,761 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation -
fs.default.name is deprecated. Instead, use fs.defaultFS
2018-07-10 23:25:09,765 [main] INFO
org.apache.pig.backend.hadoop.executionengine.HExecutionEngine - Connecting to hadoop file
system at: hdfs://localhost:8020
2018-07-10 23:25:11,103 [main] INFO  org.apache.pig.PigServer - Pig Script ID for the session:
PIG-query3.pig-fa595b35-2441-473e-82cd-6eca8f82627b
2018-07-10 23:25:11,103 [main] WARN  org.apache.pig.PigServer - ATS is disabled since
yarn.timeline-service.enabled set to false
2018-07-10 23:25:12,793 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation -
fs.default.name is deprecated. Instead, use fs.defaultFS
2018-07-10 23:25:13,049 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation -
fs.default.name is deprecated. Instead, use fs.defaultFS
2018-07-10 23:25:13,521 [main] INFO  org.apache.pig.impl.util.SpillableMemoryManager - Selected
heap (Tenured Gen) of size 699072512 to monitor. collectionUsageThreshold = 489350752,
usageThreshold = 489350752
2018-07-10 23:25:13,548 [main] INFO  org.apache.pig.tools.pigstats.ScriptState - Pig features used in
the script: HASH_JOIN,ORDER_BY,LIMIT
2018-07-10 23:25:13,638 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation -
fs.default.name is deprecated. Instead, use fs.defaultFS
2018-07-10 23:25:13,658 [main] INFO  org.apache.pig.data.SchemaTupleBackend - Key
[pig.schematuple] was not set... will not generate code.
2018-07-10 23:25:13,764 [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]}

2018-07-10 23:25:13,830 [main] INFO org.apache.pig.newplan.logical.rules.ColumnPruneVisitor - Columns pruned for EmpDetail: \$2, \$3

2018-07-10 23:25:14,079 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File concatenation threshold: 100 optimistic? false

2018-07-10 23:25:14,175 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.SecondaryKeyOptimizerMR - Using Secondary Key Optimization for MapReduce node scope-57

2018-07-10 23:25:14,203 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size before optimization: 4

2018-07-10 23:25:14,204 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size after optimization: 4

2018-07-10 23:25:14,263 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS

2018-07-10 23:25:14,320 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032

2018-07-10 23:25:14,856 [main] INFO org.apache.pig.tools.pigstats.mapreduce.MRScriptState - Pig script settings are added to the job

2018-07-10 23:25:14,878 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.reduce.markreset.buffer.percent is deprecated. Instead, use mapreduce.reduce.markreset.buffer.percent

2018-07-10 23:25:14,878 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - mapred.job.reduce.markreset.buffer.percent is not set, set to default 0.3

2018-07-10 23:25:14,878 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.output.compress is deprecated. Instead, use mapreduce.output.fileoutputformat.compress

2018-07-10 23:25:14,896 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Reduce phase detected, estimating # of required reducers.

2018-07-10 23:25:14,902 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Using reducer estimator:

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.InputSizeReducerEstimator

2018-07-10 23:25:14,960 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.InputSizeReducerEstimator - BytesPerReducer=1000000000 maxReducers=999 totalInputFileSize=333

2018-07-10 23:25:14,966 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Setting Parallelism to 1

2018-07-10 23:25:14,967 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.reduce.tasks is deprecated. Instead, use mapreduce.job.reduces

2018-07-10 23:25:14,968 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - This job cannot be converted run in-process  
2018-07-10 23:25:15,042 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.submit.replication is deprecated. Instead, use mapreduce.client.submit.file.replication  
2018-07-10 23:25:15,822 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar file:/home/acadgild/install/pig/pig-0.16.0/pig-0.16.0-core-h2.jar to DistributedCache through /tmp/temp-1449413187/tmp-223553504/pig-0.16.0-core-h2.jar  
2018-07-10 23:25:15,881 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar file:/home/acadgild/install/pig/pig-0.16.0/lib/automaton-1.11-8.jar to DistributedCache through /tmp/temp-1449413187/tmp119421004/automaton-1.11-8.jar  
2018-07-10 23:25:15,931 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar file:/home/acadgild/install/pig/pig-0.16.0/lib/antlr-runtime-3.4.jar to DistributedCache through /tmp/temp-1449413187/tmp-1077049020/antlr-runtime-3.4.jar  
2018-07-10 23:25:15,993 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar file:/home/acadgild/install/pig/pig-0.16.0/lib/joda-time-2.9.3.jar to DistributedCache through /tmp/temp-1449413187/tmp679190016/joda-time-2.9.3.jar  
2018-07-10 23:25:16,038 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Setting up single store job  
2018-07-10 23:25:16,065 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Key [pig.schematuple] is false, will not generate code.  
2018-07-10 23:25:16,065 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Starting process to move generated code to distributed cacche  
2018-07-10 23:25:16,065 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Setting key [pig.schematuple.classes] with classes to deserialize []  
2018-07-10 23:25:16,366 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 1 map-reduce job(s) waiting for submission.  
2018-07-10 23:25:16,368 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker.http.address is deprecated. Instead, use mapreduce.jobtracker.http.address  
2018-07-10 23:25:16,387 [JobControl] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032  
2018-07-10 23:25:16,436 [JobControl] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.task.id is deprecated. Instead, use mapreduce.task.attempt.id  
2018-07-10 23:25:16,440 [JobControl] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS  
2018-07-10 23:25:17,222 [JobControl] WARN org.apache.hadoop.mapreduce.JobResourceUploader - No job jar file set. User classes may not be found. See Job or Job#setJar(String).  
2018-07-10 23:25:17,344 [JobControl] INFO org.apache.pig.builtin.PigStorage - Using PigTextInputFormat  
2018-07-10 23:25:17,383 [JobControl] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1  
2018-07-10 23:25:17,384 [JobControl] INFO  
org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1  
2018-07-10 23:25:17,522 [JobControl] INFO

org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths (combined) to process : 1  
2018-07-10 23:25:17,540 [JobControl] INFO org.apache.pig.builtin.PigStorage - Using PigTextInputFormat  
2018-07-10 23:25:17,547 [JobControl] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1  
2018-07-10 23:25:17,547 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1  
2018-07-10 23:25:17,551 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths (combined) to process : 1  
2018-07-10 23:25:17,760 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter - number of splits:2  
2018-07-10 23:25:18,013 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter - Submitting tokens for job: job\_1531229668339\_0074  
2018-07-10 23:25:18,635 [JobControl] INFO org.apache.hadoop.mapred.YARNRunner - Job jar is not present. Not adding any jar to the list of resources.  
2018-07-10 23:25:18,952 [JobControl] INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl - Submitted application application\_1531229668339\_0074  
2018-07-10 23:25:19,057 [JobControl] INFO org.apache.hadoop.mapreduce.Job - The url to track the job: http://localhost:8088/proxy/application\_1531229668339\_0074/  
2018-07-10 23:25:19,057 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - HadoopJobId: job\_1531229668339\_0074  
2018-07-10 23:25:19,058 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Processing aliases EmpDetail,ExpenseDetail,commonExpense  
2018-07-10 23:25:19,058 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - detailed locations: M:  
ExpenseDetail[2,16],ExpenseDetail[-1,-1],commonExpense[3,16],EmpDetail[1,12],EmpDetail[-1,-1],commonExpense[3,16] C: R: commonExpense[3,16],commonExpense[-1,-1]  
2018-07-10 23:25:19,086 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 0% complete  
2018-07-10 23:25:19,086 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Running jobs are [job\_1531229668339\_0074]  
ss : 1  
2018-07-10 23:25:17,384 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1  
2018-07-10 23:25:17,522 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths (combined) to process : 1  
2018-07-10 23:25:17,540 [JobControl] INFO org.apache.pig.builtin.PigStorage - Using PigTextInputFormat  
2018-07-10 23:25:17,547 [JobControl] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1  
2018-07-10 23:25:17,547 [JobControl] INFO

org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1  
2018-07-10 23:25:17,551 [JobControl] INFO  
org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths (combined) to process : 1  
2018-07-10 23:25:17,760 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter - number of splits:2  
2018-07-10 23:25:18,013 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter - Submitting tokens for job: job\_1531229668339\_0074  
2018-07-10 23:25:18,635 [JobControl] INFO org.apache.hadoop.mapred.YARNRunner - Job jar is not present. Not adding any jar to the list of resources.  
2018-07-10 23:25:18,952 [JobControl] INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl - Submitted application application\_1531229668339\_0074  
2018-07-10 23:25:19,057 [JobControl] INFO org.apache.hadoop.mapreduce.Job - The url to track the job: http://localhost:8088/proxy/application\_1531229668339\_0074/  
2018-07-10 23:25:19,057 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - HadoopJobId: job\_1531229668339\_0074  
2018-07-10 23:25:19,058 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Processing aliases EmpDetail,ExpenseDetail,commonExpense  
2018-07-10 23:25:19,058 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - detailed locations: M:  
ExpenseDetail[2,16],ExpenseDetail[-1,-1],commonExpense[3,16],EmpDetail[1,12],EmpDetail[-1,-1],commonExpense[3,16] C: R: commonExpense[3,16],commonExpense[-1,-1]  
2018-07-10 23:25:19,086 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 0% complete  
2018-07-10 23:25:19,086 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Running jobs are [job\_1531229668339\_0074]  
2018-07-10 23:25:49,402 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 6% complete  
2018-07-10 23:25:49,402 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Running jobs are [job\_1531229668339\_0074]  
2018-07-10 23:25:51,409 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 12% complete  
2018-07-10 23:25:51,409 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Running jobs are [job\_1531229668339\_0074]  
2018-07-10 23:26:03,451 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 25% complete  
2018-07-10 23:26:03,452 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Running jobs are [job\_1531229668339\_0074]



2018-07-10 23:26:04,466 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032

2018-07-10 23:26:04,503 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

2018-07-10 23:26:05,335 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032

2018-07-10 23:26:05,354 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

2018-07-10 23:26:05,484 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032

2018-07-10 23:26:05,492 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

2018-07-10 23:26:05,599 [main] INFO org.apache.pig.tools.pigstats.mapreduce.MRScriptState - Pig script settings are added to the job

2018-07-10 23:26:05,600 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - mapred.job.reduce.markreset.buffer.percent is not set, set to default 0.3

2018-07-10 23:26:05,602 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Reduce phase detected, estimating # of required reducers.

2018-07-10 23:26:05,602 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Using reducer estimator:

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.InputSizeReducerEstimator

2018-07-10 23:26:05,631 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.InputSizeReducerEstimator - BytesPerReducer=1000000000 maxReducers=999 totalInputFileSize=159

2018-07-10 23:26:05,633 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Setting Parallelism to 1

2018-07-10 23:26:05,633 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - This job cannot be converted run in-process

2018-07-10 23:26:05,861 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar file:/home/acadgild/install/pig/pig-0.16.0/pig-0.16.0-core-h2.jar to DistributedCache through /tmp/temp-1449413187/tmp1853846880/pig-0.16.0-core-h2.jar

2018-07-10 23:26:05,937 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar file:/home/acadgild/install/pig/pig-0.16.0/lib/automaton-1.11-8.jar to DistributedCache through /tmp/temp-1449413187/tmp-2109604730/automaton-1.11-8.jar

2018-07-10 23:26:05,971 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar file:/home/acadgild/install/pig/pig-0.16.0/lib/antlr-runtime-3.4.jar to DistributedCache through /tmp/temp-1449413187/tmp784113131/antlr-runtime-3.4.jar

2018-07-10 23:26:06,059 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar file:/home/acadgild/install/pig/pig-0.16.0/lib/joda-time-2.9.3.jar to DistributedCache through /tmp/temp-1449413187/tmp1185326890/joda-time-2.9.3.jar  
2018-07-10 23:26:06,061 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Setting up single store job  
2018-07-10 23:26:06,064 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Key [pig.schematuple] is false, will not generate code.  
2018-07-10 23:26:06,064 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Starting process to move generated code to distributed cacche  
2018-07-10 23:26:06,064 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Setting key [pig.schematuple.classes] with classes to deserialize []  
2018-07-10 23:26:06,104 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 1 map-reduce job(s) waiting for submission.  
2018-07-10 23:26:06,112 [JobControl] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032  
2018-07-10 23:26:06,133 [JobControl] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS  
2018-07-10 23:26:06,159 [JobControl] WARN org.apache.hadoop.mapreduce.JobResourceUploader - No job jar file set. User classes may not be found. See Job or Job#setJar(String).  
2018-07-10 23:26:06,211 [JobControl] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1  
2018-07-10 23:26:06,212 [JobControl] INFO  
org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1  
2018-07-10 23:26:06,212 [JobControl] INFO  
org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths (combined) to process : 1  
2018-07-10 23:26:06,375 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter - number of splits:1  
2018-07-10 23:26:06,453 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter - Submitting tokens for job: job\_1531229668339\_0075  
2018-07-10 23:26:06,459 [JobControl] INFO org.apache.hadoop.mapred.YARNRunner - Job jar is not present. Not adding any jar to the list of resources.  
2018-07-10 23:26:06,527 [JobControl] INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl - Submitted application application\_1531229668339\_0075  
2018-07-10 23:26:06,535 [JobControl] INFO org.apache.hadoop.mapreduce.Job - The url to track the job: http://localhost:8088/proxy/application\_1531229668339\_0075/  
2018-07-10 23:26:06,607 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - HadoopJobId: job\_1531229668339\_0075  
2018-07-10 23:26:06,607 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Processing aliases expensesOrder  
2018-07-10 23:26:06,607 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - detailed locations: M: expensesOrder[4,16] C: R:  
2018-07-10 23:26:33,927 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 37%

complete

2018-07-10 23:26:33,927 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Running jobs are [job\_1531229668339\_0075]

2018-07-10 23:26:53,987 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 50% complete

2018-07-10 23:26:53,987 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Running jobs are [job\_1531229668339\_0075]

2018-07-10 23:26:56,999 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032

2018-07-10 23:26:57,013 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

2018-07-10 23:26:57,306 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032

2018-07-10 23:26:57,328 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

2018-07-10 23:26:57,433 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032

2018-07-10 23:26:57,456 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

2018-07-10 23:26:57,585 [main] INFO org.apache.pig.tools.pigstats.mapreduce.MRScriptState - Pig script settings are added to the job

2018-07-10 23:26:57,585 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - mapred.job.reduce.markreset.buffer.percent is not set, set to default 0.3

2018-07-10 23:26:57,586 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Reduce phase detected, estimating # of required reducers.

2018-07-10 23:26:57,586 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Setting Parallelism to 1

2018-07-10 23:26:57,586 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - This job cannot be converted run in-process

2018-07-10 23:26:57,796 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar file:/home/acadgild/install/pig/pig-0.16.0/pig-0.16.0-core-h2.jar to DistributedCache through /tmp/temp-1449413187/tmp-1278186776/pig-0.16.0-core-h2.jar

2018-07-10 23:26:57,848 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar file:/home/acadgild/install/pig/pig-0.16.0/lib/automaton-1.11-8.jar to DistributedCache through /tmp/temp-1449413187/tmp1925723372/automaton-1.11-8.jar

2018-07-10 23:26:57,903 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar

file:/home/acadgild/install/pig/pig-0.16.0/lib/antlr-runtime-3.4.jar to DistributedCache through  
/tmp/temp-1449413187/tmp1638821155/antlr-runtime-3.4.jar  
2018-07-10 23:26:57,973 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar  
file:/home/acadgild/install/pig/pig-0.16.0/lib/joda-time-2.9.3.jar to DistributedCache through  
/tmp/temp-1449413187/tmp-908825641/joda-time-2.9.3.jar  
2018-07-10 23:26:58,016 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Setting up  
single store job  
2018-07-10 23:26:58,022 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Key  
[pig.schematuple] is false, will not generate code.  
2018-07-10 23:26:58,022 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Starting process  
to move generated code to distributed cacche  
2018-07-10 23:26:58,022 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Setting key  
[pig.schematuple.classes] with classes to deserialize []  
2018-07-10 23:26:58,153 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 1  
map-reduce job(s) waiting for submission.  
2018-07-10 23:26:58,156 [JobControl] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to  
ResourceManager at localhost/127.0.0.1:8032  
2018-07-10 23:26:58,166 [JobControl] INFO org.apache.hadoop.conf.Configuration.deprecation -  
fs.default.name is deprecated. Instead, use fs.defaultFS  
2018-07-10 23:26:58,220 [JobControl] WARN org.apache.hadoop.mapreduce.JobResourceUploader -  
No job jar file set. User classes may not be found. See Job or Job#setJar(String).  
2018-07-10 23:26:58,285 [JobControl] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat  
- Total input paths to process : 1  
2018-07-10 23:26:58,286 [JobControl] INFO  
org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1  
2018-07-10 23:26:58,289 [JobControl] INFO  
org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths (combined) to  
process : 1  
2018-07-10 23:26:58,407 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter - number of  
splits:1  
2018-07-10 23:26:58,501 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter -  
Submitting tokens for job: job\_1531229668339\_0076  
2018-07-10 23:26:58,511 [JobControl] INFO org.apache.hadoop.mapred.YARNRunner - Job jar is not  
present. Not adding any jar to the list of resources.  
2018-07-10 23:26:58,627 [JobControl] INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl -  
Submitted application application\_1531229668339\_0076  
2018-07-10 23:26:58,632 [JobControl] INFO org.apache.hadoop.mapreduce.Job - The url to track the  
job: http://localhost:8088/proxy/application\_1531229668339\_0076/  
2018-07-10 23:26:58,655 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher -  
HadoopJobId: job\_1531229668339\_0076  
2018-07-10 23:26:58,655 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Processing  
aliases expensesOrder  
2018-07-10 23:26:58,656 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - detailed

locations: M: expensesOrder[4,16] C: R:  
2018-07-10 23:27:30,992 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 62%  
complete  
2018-07-10 23:27:30,992 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Running  
jobs are [job\_1531229668339\_0076]  
2018-07-10 23:27:43,027 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 75%  
complete  
2018-07-10 23:27:43,027 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Running  
jobs are [job\_1531229668339\_0076]  
2018-07-10 23:27:44,049 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to  
ResourceManager at localhost/127.0.0.1:8032  
2018-07-10 23:27:44,073 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate -  
Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history  
server  
2018-07-10 23:27:44,236 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to  
ResourceManager at localhost/127.0.0.1:8032  
2018-07-10 23:27:44,242 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate -  
Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history  
server  
2018-07-10 23:27:44,328 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to  
ResourceManager at localhost/127.0.0.1:8032  
2018-07-10 23:27:44,334 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate -  
Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history  
server  
2018-07-10 23:27:44,402 [main] INFO org.apache.pig.tools.pigstats.mapreduce.MRScriptState - Pig  
script settings are added to the job  
2018-07-10 23:27:44,402 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler -  
mapred.job.reduce.markreset.buffer.percent is not set, set to default 0.3  
2018-07-10 23:27:44,403 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Reduce phase  
detected, estimating # of required reducers.  
2018-07-10 23:27:44,403 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Setting  
Parallelism to 1  
2018-07-10 23:27:44,403 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - This job  
cannot be converted run in-process  
2018-07-10 23:27:44,527 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar  
file:/home/acadgild/install/pig/pig-0.16.0/pig-0.16.0-core-h2.jar to DistributedCache through  
/tmp/temp-1449413187/tmp2005539124/pig-0.16.0-core-h2.jar  
2018-07-10 23:27:44,582 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar  
file:/home/acadgild/install/pig/pig-0.16.0/lib/automaton-1.11-8.jar to DistributedCache through

/tmp/temp-1449413187/tmp516128024/automaton-1.11-8.jar  
2018-07-10 23:27:44,650 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar  
file:/home/acadgild/install/pig/pig-0.16.0/lib/antlr-runtime-3.4.jar to DistributedCache through  
/tmp/temp-1449413187/tmp-1213406232/antlr-runtime-3.4.jar  
2018-07-10 23:27:44,738 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar  
file:/home/acadgild/install/pig/pig-0.16.0/lib/joda-time-2.9.3.jar to DistributedCache through  
/tmp/temp-1449413187/tmp1691894694/joda-time-2.9.3.jar  
2018-07-10 23:27:44,741 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Setting up  
single store job  
2018-07-10 23:27:44,743 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Key  
[pig.schematuple] is false, will not generate code.  
2018-07-10 23:27:44,743 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Starting process  
to move generated code to distributed cacche  
2018-07-10 23:27:44,743 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Setting key  
[pig.schematuple.classes] with classes to deserialize []  
2018-07-10 23:27:44,769 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 1  
map-reduce job(s) waiting for submission.  
2018-07-10 23:27:44,771 [JobControl] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to  
ResourceManager at localhost/127.0.0.1:8032  
2018-07-10 23:27:44,782 [JobControl] INFO org.apache.hadoop.conf.Configuration.deprecation -  
fs.default.name is deprecated. Instead, use fs.defaultFS  
2018-07-10 23:27:44,801 [JobControl] WARN org.apache.hadoop.mapreduce.JobResourceUploader -  
No job jar file set. User classes may not be found. See Job or Job#setJar(String).  
2018-07-10 23:27:44,831 [JobControl] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat  
- Total input paths to process : 1  
2018-07-10 23:27:44,831 [JobControl] INFO  
org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1  
2018-07-10 23:27:44,832 [JobControl] INFO  
org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths (combined) to  
process : 1  
2018-07-10 23:27:44,935 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter - number of  
splits:1  
2018-07-10 23:27:44,987 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter -  
Submitting tokens for job: job\_1531229668339\_0077  
2018-07-10 23:27:44,994 [JobControl] INFO org.apache.hadoop.mapred.YARNRunner - Job jar is not  
present. Not adding any jar to the list of resources.  
2018-07-10 23:27:45,069 [JobControl] INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl -  
Submitted application application\_1531229668339\_0077  
2018-07-10 23:27:45,078 [JobControl] INFO org.apache.hadoop.mapreduce.Job - The url to track the  
job: http://localhost:8088/proxy/application\_1531229668339\_0077/  
2018-07-10 23:27:45,270 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher -  
HadoopJobId: job\_1531229668339\_0077  
2018-07-10 23:27:45,270 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Processing

```

aliases expenseMaxEmp,expensesOrder
2018-07-10 23:27:45,270 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - detailed
locations: M: expensesOrder[4,16] C: R: expensesOrder[-1,-1],expenseMaxEmp[6,16]
2018-07-10 23:28:12,463 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 87%
complete
2018-07-10 23:28:12,464 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Running
jobs are [job_1531229668339_0077]
2018-07-10 23:28:22,493 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Running
jobs are [job_1531229668339_0077]
2018-07-10 23:28:25,504 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to
ResourceManager at localhost/127.0.0.1:8032
2018-07-10 23:28:25,527 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate -
Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history
server
2018-07-10 23:28:25,826 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to
ResourceManager at localhost/127.0.0.1:8032
2018-07-10 23:28:25,842 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate -
Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history
server
2018-07-10 23:28:26,022 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to
ResourceManager at localhost/127.0.0.1:8032
2018-07-10 23:28:26,048 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate -
Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history
server
2018-07-10 23:28:26,154 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 100%
complete
2018-07-10 23:28:26,233 [main] INFO org.apache.pig.tools.pigstats.mapreduce.SimplePigStats -
Script Statistics:

```

```

HadoopVersion      PigVersion  UserId StartedAt   FinishedAt  Features
2.6.5  0.16.0 acadgild    2018-07-10 23:25:14 2018-07-10 23:28:26
                HASH_JOIN,ORDER_BY,LIMIT

```

Success!

Job Stats (time in seconds):

JobId	Maps	Reduces	MaxMapTime	MinMapTime	AvgMapTime	MedianMapTime	MaxRe
duceTime	MinReduceTime	AvgReduceTime	MedianReductime	Alias	Feature		
Outputs							
job_1531229668339_0074	2	1	14	14	14	10	10
EmpDetail,ExpenseDetail,commonExpense							
job_1531229668339_0075	1	1	10	10	10	15	15
expensesOrder							
job_1531229668339_0076	1	1	8	8	8	9	9

expensesOrder ORDER\_BY,COMBINER  
job\_1531229668339\_0077 1 1 7 7 7 7 8 8 8 8  
expenseMaxEmp,expensesOrder hdfs://localhost:8020/tmp/temp-1449413187/tmp1158159175,  
75,

Input(s):

Successfully read 14 records from: "/user/acadgild/hadoopdata/pig/input/employee\_details.txt"  
Successfully read 9 records from: "/user/acadgild/hadoopdata/pig/input/employee\_expenses.txt"

Output(s):

Successfully stored 1 records (17 bytes) in:  
"hdfs://localhost:8020/tmp/temp-1449413187/tmp1158159175"

Counters:

Total records written : 1  
Total bytes written : 17  
Spillable Memory Manager spill count : 0  
Total bags proactively spilled: 0  
Total records proactively spilled: 0

Job DAG:

job\_1531229668339\_0074 -> job\_1531229668339\_0075,  
job\_1531229668339\_0075 -> job\_1531229668339\_0076,  
job\_1531229668339\_0076 -> job\_1531229668339\_0077,  
job\_1531229668339\_0077

2018-07-10 23:28:26,235 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032

2018-07-10 23:28:26,241 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

2018-07-10 23:28:26,364 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032

2018-07-10 23:28:26,374 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

2018-07-10 23:28:26,486 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032

2018-07-10 23:28:26,497 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

2018-07-10 23:28:26,604 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032

2018-07-10 23:28:26,609 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

2018-07-10 23:28:26,692 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032



2018-07-10 23:28:26,698 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

2018-07-10 23:28:26,774 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032

2018-07-10 23:28:26,785 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

2018-07-10 23:28:26,866 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032

2018-07-10 23:28:26,872 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

2018-07-10 23:28:26,970 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032

2018-07-10 23:28:26,975 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

2018-07-10 23:28:27,058 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032

2018-07-10 23:28:27,063 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

2018-07-10 23:28:27,137 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032

2018-07-10 23:28:27,142 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

2018-07-10 23:28:27,213 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032

2018-07-10 23:28:27,218 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

2018-07-10 23:28:27,282 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032

2018-07-10 23:28:27,288 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

2018-07-10 23:28:27,360 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!

2018-07-10 23:28:27,367 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS

2018-07-10 23:28:27,369 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.

2018-07-10 23:28:27,380 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1

2018-07-10 23:28:27,381 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1

**(110,Priyanka)**

2018-07-10 23:28:27,569 [main] INFO org.apache.pig.Main - Pig script completed in 3 minutes, 20 seconds and 201 milliseconds (200201 ms)

**(d) List of employees (employee id and employee name) having entries in employee\_expenses file.**

**Pig Script : filename: query4.pig**

```
EmpDetail = LOAD '/user/acadgild/hadoopdata/pig/input/employee_details.txt' USING PigStorage(',')
AS(EmpID:int,name:chararray,salary:int,DepartmentID:int);
ExpenseDetail = LOAD '/user/acadgild/hadoopdata/pig/input/employee_expenses.txt' USING
PigStorage(' ') AS (id:int,expense:int);
commonExpense = JOIN EmpDetail by eid RIGHT, ExpenseDetail by id;
listExpense = FOREACH commonExpense GENERATE id,name;
distinct1 = DISTINCT listExpense;
dump distinct1;
```

**Terminal Execution:**

```
[acadgild@localhost ~]$ pig /home/acadgild/Desktop/query4.pig
18/07/10 23:38:05 INFO pig.ExecTypeProvider: Trying ExecType : LOCAL
18/07/10 23:38:05 INFO pig.ExecTypeProvider: Trying ExecType : MAPREDUCE
18/07/10 23:38:05 INFO pig.ExecTypeProvider: Picked MAPREDUCE as the ExecType
2018-07-10 23:38:05,982 [main] INFO org.apache.pig.Main - Apache Pig version 0.16.0 (r1746530)
compiled Jun 01 2016, 23:10:49
2018-07-10 23:38:05,982 [main] INFO org.apache.pig.Main - Logging error messages to:
/home/acadgild/pig_1531246085979.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-1.4.4/lib/slf4j-log4j12-1.7.10.jar!/org/slf4j/impl/StaticLogg
erBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
2018-07-10 23:38:06,853 [main] WARN org.apache.hadoop.util.NativeCodeLoader - Unable to load
native-hadoop library for your platform... using builtin-java classes where applicable
2018-07-10 23:38:07,235 [main] INFO org.apache.pig.impl.util.Utils - Default bootup file
/home/acadgild/.pigbootup not found
2018-07-10 23:38:07,433 [main] INFO org.apache.hadoop.conf.Configuration.deprecation -
```

mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address  
2018-07-10 23:38:07,433 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS  
2018-07-10 23:38:07,436 [main] INFO org.apache.pig.backend.hadoop.executionengine.HExecutionEngine - Connecting to hadoop file system at: hdfs://localhost:8020  
2018-07-10 23:38:08,497 [main] INFO org.apache.pig.PigServer - Pig Script ID for the session: PIG-query4.pig-b26a78f7-fbc2-4682-877f-00a1a34f999e  
2018-07-10 23:38:08,497 [main] WARN org.apache.pig.PigServer - ATS is disabled since yarn.timeline-service.enabled set to false  
2018-07-10 23:38:09,855 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS  
2018-07-10 23:38:10,015 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS  
2018-07-10 23:38:10,438 [main] INFO org.apache.pig.impl.util.SpillableMemoryManager - Selected heap (Tenured Gen) of size 699072512 to monitor. collectionUsageThreshold = 489350752, usageThreshold = 489350752  
2018-07-10 23:38:10,483 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: HASH\_JOIN,DISTINCT  
2018-07-10 23:38:10,580 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS  
2018-07-10 23:38:10,604 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.  
2018-07-10 23:38:10,700 [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]}  
2018-07-10 23:38:10,790 [main] INFO org.apache.pig.newplan.logical.rules.ColumnPruneVisitor - Columns pruned for EmpDetail: \$2, \$3  
2018-07-10 23:38:10,797 [main] INFO org.apache.pig.newplan.logical.rules.ColumnPruneVisitor - Columns pruned for ExpenseDetail: \$1  
2018-07-10 23:38:11,059 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File concatenation threshold: 100 optimistic? false  
2018-07-10 23:38:11,181 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size before optimization: 2  
2018-07-10 23:38:11,183 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size after optimization: 2  
2018-07-10 23:38:11,270 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS  
2018-07-10 23:38:11,352 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032  
2018-07-10 23:38:11,804 [main] INFO org.apache.pig.tools.pigstats.mapreduce.MRScriptState - Pig script settings are added to the job  
2018-07-10 23:38:11,828 [main] INFO org.apache.hadoop.conf.Configuration.deprecation -

mapred.job.reduce.markreset.buffer.percent is deprecated. Instead, use  
mapreduce.reduce.markreset.buffer.percent  
2018-07-10 23:38:11,828 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler -  
mapred.job.reduce.markreset.buffer.percent is not set, set to default 0.3  
2018-07-10 23:38:11,829 [main] INFO org.apache.hadoop.conf.Configuration.deprecation -  
mapred.output.compress is deprecated. Instead, use mapreduce.output.fileoutputformat.compress  
2018-07-10 23:38:11,837 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Reduce phase  
detected, estimating # of required reducers.  
2018-07-10 23:38:11,846 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Using  
reducer estimator:  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.InputSizeReducerEstimator  
2018-07-10 23:38:11,890 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.InputSizeReducerEstimator -  
BytesPerReducer=1000000000 maxReducers=999 totalInputFileSize=333  
2018-07-10 23:38:11,891 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Setting  
Parallelism to 1  
2018-07-10 23:38:11,891 [main] INFO org.apache.hadoop.conf.Configuration.deprecation -  
mapred.reduce.tasks is deprecated. Instead, use mapreduce.job.reduces  
2018-07-10 23:38:11,894 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - This job  
cannot be converted run in-process  
2018-07-10 23:38:11,945 [main] INFO org.apache.hadoop.conf.Configuration.deprecation -  
mapred.submit.replication is deprecated. Instead, use mapreduce.client.submit.file.replication  
2018-07-10 23:38:12,546 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar  
file:/home/acadgild/install/pig/pig-0.16.0/pig-0.16.0-core-h2.jar to DistributedCache through  
/tmp/temp-1231858679/tmp1829445756/pig-0.16.0-core-h2.jar  
2018-07-10 23:38:12,602 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar  
file:/home/acadgild/install/pig/pig-0.16.0/lib/automaton-1.11-8.jar to DistributedCache through  
/tmp/temp-1231858679/tmp963026660/automaton-1.11-8.jar  
2018-07-10 23:38:12,762 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar  
file:/home/acadgild/install/pig/pig-0.16.0/lib/antlr-runtime-3.4.jar to DistributedCache through  
/tmp/temp-1231858679/tmp-140619027/antlr-runtime-3.4.jar  
2018-07-10 23:38:12,858 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar  
file:/home/acadgild/install/pig/pig-0.16.0/lib/joda-time-2.9.3.jar to DistributedCache through  
/tmp/temp-1231858679/tmp531424095/joda-time-2.9.3.jar  
2018-07-10 23:38:12,896 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Setting up  
single store job  
2018-07-10 23:38:12,949 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Key  
[pig.schematuple] is false, will not generate code.  
2018-07-10 23:38:12,949 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Starting process

to move generated code to distributed cache

2018-07-10 23:38:12,950 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Setting key [pig.schematuple.classes] with classes to deserialize []

2018-07-10 23:38:13,189 [main] INFO

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

2018-07-10 23:38:13,191 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker.http.address is deprecated. Instead, use mapreduce.jobtracker.http.address

2018-07-10 23:38:13,197 [JobControl] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032

2018-07-10 23:38:13,246 [JobControl] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.task.id is deprecated. Instead, use mapreduce.task.attempt.id

2018-07-10 23:38:13,251 [JobControl] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS

2018-07-10 23:38:13,814 [JobControl] WARN org.apache.hadoop.mapreduce.JobResourceUploader - No job jar file set. User classes may not be found. See Job or Job#setJar(String).

2018-07-10 23:38:13,889 [JobControl] INFO org.apache.pig.builtin.PigStorage - Using PigTextInputFormat

2018-07-10 23:38:13,911 [JobControl] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1

2018-07-10 23:38:13,911 [JobControl] INFO

org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1

2018-07-10 23:38:13,995 [JobControl] INFO

org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths (combined) to process : 1

2018-07-10 23:38:14,000 [JobControl] INFO org.apache.pig.builtin.PigStorage - Using PigTextInputFormat

2018-07-10 23:38:14,004 [JobControl] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1

2018-07-10 23:38:14,004 [JobControl] INFO

org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1

2018-07-10 23:38:14,007 [JobControl] INFO

org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths (combined) to process : 1

2018-07-10 23:38:14,195 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter - number of splits:2

2018-07-10 23:38:14,481 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter - Submitting tokens for job: job\_1531229668339\_0079

2018-07-10 23:38:14,815 [JobControl] INFO org.apache.hadoop.mapred.YARNRunner - Job jar is not present. Not adding any jar to the list of resources.

2018-07-10 23:38:15,051 [JobControl] INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl - Submitted application application\_1531229668339\_0079

2018-07-10 23:38:15,205 [JobControl] INFO org.apache.hadoop.mapreduce.Job - The url to track the job: http://localhost:8088/proxy/application\_1531229668339\_0079/

2018-07-10 23:38:15,205 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - HadoopJobId: job\_1531229668339\_0079

2018-07-10 23:38:15,205 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Processing

aliases EmpDetail,ExpenseDetail,commonExpense,listExpense  
2018-07-10 23:38:15,205 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - detailed  
locations: M:  
ExpenseDetail[2,16],ExpenseDetail[-1,-1],commonExpense[4,16],EmpDetail[1,12],EmpDetail[-1,-1],c  
ommonExpense[4,16] C: R: commonExpense[4,16],listExpense[5,14]  
2018-07-10 23:38:15,242 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 0%  
complete  
2018-07-10 23:38:15,243 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Running  
jobs are [job\_1531229668339\_0079]  
2018-07-10 23:38:47,585 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 25%  
complete  
2018-07-10 23:38:47,585 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Running  
jobs are [job\_1531229668339\_0079]  
2018-07-10 23:38:59,618 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 50%  
complete  
2018-07-10 23:38:59,618 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Running  
jobs are [job\_1531229668339\_0079]  
2018-07-10 23:39:00,631 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to  
ResourceManager at localhost/127.0.0.1:8032  
2018-07-10 23:39:00,658 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate -  
Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history  
server  
2018-07-10 23:39:01,055 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to  
ResourceManager at localhost/127.0.0.1:8032  
2018-07-10 23:39:01,066 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate -  
Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history  
server  
2018-07-10 23:39:01,234 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to  
ResourceManager at localhost/127.0.0.1:8032  
2018-07-10 23:39:01,242 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate -  
Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history  
server  
2018-07-10 23:39:01,432 [main] INFO org.apache.pig.tools.pigstats.mapreduce.MRScriptState - Pig  
script settings are added to the job  
2018-07-10 23:39:01,432 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler -  
mapred.job.reduce.markreset.buffer.percent is not set, set to default 0.3  
2018-07-10 23:39:01,433 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Reduce phase  
detected, estimating # of required reducers.  
2018-07-10 23:39:01,433 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Using

reducer estimator:

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.InputSizeReducerEstimator  
2018-07-10 23:39:01,474 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.InputSizeReducerEstimator -  
BytesPerReducer=1000000000 maxReducers=999 totalInputFileSize=136

2018-07-10 23:39:01,475 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Setting  
Parallelism to 1

2018-07-10 23:39:01,475 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - This job  
cannot be converted run in-process

2018-07-10 23:39:01,643 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar  
file:/home/acadgild/install/pig/pig-0.16.0/pig-0.16.0-core-h2.jar to DistributedCache through  
/tmp/temp-1231858679/tmp-1750242937/pig-0.16.0-core-h2.jar

2018-07-10 23:39:01,720 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar  
file:/home/acadgild/install/pig/pig-0.16.0/lib/automaton-1.11-8.jar to DistributedCache through  
/tmp/temp-1231858679/tmp1225807687/automaton-1.11-8.jar

2018-07-10 23:39:01,812 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar  
file:/home/acadgild/install/pig/pig-0.16.0/lib/antlr-runtime-3.4.jar to DistributedCache through  
/tmp/temp-1231858679/tmp515328088/antlr-runtime-3.4.jar

2018-07-10 23:39:01,922 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar  
file:/home/acadgild/install/pig/pig-0.16.0/lib/joda-time-2.9.3.jar to DistributedCache through  
/tmp/temp-1231858679/tmp-947235839/joda-time-2.9.3.jar

2018-07-10 23:39:01,928 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Setting up  
single store job

2018-07-10 23:39:01,929 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Key  
[pig.schematuple] is false, will not generate code.

2018-07-10 23:39:01,929 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Starting process  
to move generated code to distributed cacche

2018-07-10 23:39:01,929 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Setting key  
[pig.schematuple.classes] with classes to deserialize []

2018-07-10 23:39:01,932 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Setting  
identity combiner class.

2018-07-10 23:39:01,983 [main] INFO

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

2018-07-10 23:39:01,987 [JobControl] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to  
ResourceManager at localhost/127.0.0.1:8032

2018-07-10 23:39:02,007 [JobControl] INFO org.apache.hadoop.conf.Configuration.deprecation -  
fs.default.name is deprecated. Instead, use fs.defaultFS

2018-07-10 23:39:02,039 [JobControl] WARN org.apache.hadoop.mapreduce.JobResourceUploader -  
No job jar file set. User classes may not be found. See Job or Job#setJar(String).

2018-07-10 23:39:02,102 [JobControl] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat

- Total input paths to process : 1  
2018-07-10 23:39:02,102 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1  
2018-07-10 23:39:02,102 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths (combined) to process : 1  
2018-07-10 23:39:02,218 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter - number of splits:1  
2018-07-10 23:39:02,315 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter - Submitting tokens for job: job\_1531229668339\_0080  
2018-07-10 23:39:02,324 [JobControl] INFO org.apache.hadoop.mapred.YARNRunner - Job jar is not present. Not adding any jar to the list of resources.  
2018-07-10 23:39:02,405 [JobControl] INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl - Submitted application application\_1531229668339\_0080  
2018-07-10 23:39:02,424 [JobControl] INFO org.apache.hadoop.mapreduce.Job - The url to track the job: http://localhost:8088/proxy/application\_1531229668339\_0080/  
2018-07-10 23:39:02,485 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - HadoopJobId: job\_1531229668339\_0080  
2018-07-10 23:39:02,485 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Processing aliases  
2018-07-10 23:39:02,485 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - detailed locations: M: C: R:  
2018-07-10 23:39:29,692 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 75% complete  
2018-07-10 23:39:29,692 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Running jobs are [job\_1531229668339\_0080]  
2018-07-10 23:39:49,752 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Running jobs are [job\_1531229668339\_0080]  
2018-07-10 23:39:52,774 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032  
2018-07-10 23:39:52,813 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server  
2018-07-10 23:39:53,287 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032  
2018-07-10 23:39:53,339 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server  
2018-07-10 23:39:53,573 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032  
2018-07-10 23:39:53,592 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server



2018-07-10 23:39:53,792 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 100% complete  
2018-07-10 23:39:53,910 [main] INFO org.apache.pig.tools.pigstats.mapreduce.SimplePigStats - Script Statistics:

HadoopVersion	PigVersion	UserId	StartedAt	FinishedAt	Features
2.6.5	0.16.0	acadgild	2018-07-10 23:38:11	2018-07-10 23:39:53	HASH_JOIN,DISTINCT

Success!

Job Stats (time in seconds):

JobId	Maps	Reduces	MaxMapTime	MinMapTime	AvgMapTime	MedianMapTime	MaxReduceTime	MinReduceTime	AvgReduceTime	MedianReduceTime	Alias	Feature
Outputs												
job_1531229668339_0079	2	1	16	16	16	16	10	10	10	10	10	10
EmpDetail,ExpenseDetail,commonExpense,listExpense										HASH_JOIN		
job_1531229668339_0080	1	1	9	9	9	9	17	17	17	17	17	17
DISTINCT hdfs://localhost:8020/tmp/temp-1231858679/tmp-651752057,												

Input(s):

Successfully read 9 records from: "/user/acadgild/hadoopdata/pig/input/employee\_expenses.txt"  
Successfully read 14 records from: "/user/acadgild/hadoopdata/pig/input/employee\_details.txt"

Output(s):

Successfully stored 7 records (103 bytes) in:  
"hdfs://localhost:8020/tmp/temp-1231858679/tmp-651752057"

Counters:

Total records written : 7  
Total bytes written : 103  
Spillable Memory Manager spill count : 0  
Total bags proactively spilled: 0  
Total records proactively spilled: 0

Job DAG:

job\_1531229668339\_0079 -> job\_1531229668339\_0080,  
job\_1531229668339\_0080

2018-07-10 23:39:53,929 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032

2018-07-10 23:39:53,947 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

2018-07-10 23:39:54,325 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032

2018-07-10 23:39:54,347 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

server

2018-07-10 23:39:54,920 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032

2018-07-10 23:39:54,960 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

2018-07-10 23:39:55,321 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032

2018-07-10 23:39:55,338 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

2018-07-10 23:39:55,503 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032

2018-07-10 23:39:55,523 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

2018-07-10 23:39:55,692 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032

2018-07-10 23:39:55,702 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

2018-07-10 23:39:55,835 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!

2018-07-10 23:39:55,856 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS

2018-07-10 23:39:55,858 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.

2018-07-10 23:39:55,877 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1

2018-07-10 23:39:55,877 [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)**

**(119,)**

2018-07-10 23:39:56,377 [main] INFO org.apache.pig.Main - Pig script completed in 1 minute, 50 seconds and 743 milliseconds (110743 ms)

[acadgild@localhost ~]\$

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

**Pig Script : filename: query5.pig**

EmpDetail = LOAD '/user/acadgild/hadoopdata/pig/input/employee\_details.txt' USING PigStorage(',')

```

AS(EmpID:int,name:chararray,salary:int,DepartmentID:int);
ExpenseDetail = LOAD '/user/acadgild/hadoopdata/pig/input/employee_expenses.txt' USING
PigStorage(' ') AS (id:int,expense:int);
nonCommonEmp = JOIN EmpDetail by EmpID LEFT, ExpenseDetail by id;
uncommon1 = FILTER nonCommonEmp BY id is null;
Sol5 = FOREACH uncommon1 GENERATE EmpID,name;
dump Sol5;

```

### **Terminal Execution:**

```

[acadgild@localhost ~]$ pig /home/acadgild/Desktop/query5.pig
18/07/10 23:32:51 INFO pig.ExecTypeProvider: Trying ExecType : LOCAL
18/07/10 23:32:51 INFO pig.ExecTypeProvider: Trying ExecType : MAPREDUCE
18/07/10 23:32:51 INFO pig.ExecTypeProvider: Picked MAPREDUCE as the ExecType
2018-07-10 23:32:51,771 [main] INFO org.apache.pig.Main - Apache Pig version 0.16.0 (r1746530)
compiled Jun 01 2016, 23:10:49
2018-07-10 23:32:51,772 [main] INFO org.apache.pig.Main - Logging error messages to:
/home/acadgild/pig_1531245771768.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-1.4.4/lib/slf4j-log4j12-1.7.10.jar!/org/slf4j/impl/StaticLogg
erBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
2018-07-10 23:32:52,662 [main] WARN org.apache.hadoop.util.NativeCodeLoader - Unable to load
native-hadoop library for your platform... using builtin-java classes where applicable
2018-07-10 23:32:53,049 [main] INFO org.apache.pig.impl.util.Utils - Default bootup file
/home/acadgild/.pigbootup not found
2018-07-10 23:32:53,246 [main] INFO org.apache.hadoop.conf.Configuration.deprecation -
mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2018-07-10 23:32:53,246 [main] INFO org.apache.hadoop.conf.Configuration.deprecation -
fs.default.name is deprecated. Instead, use fs.defaultFS
2018-07-10 23:32:53,250 [main] INFO
org.apache.pig.backend.hadoop.executionengine.HExecutionEngine - Connecting to hadoop file
system at: hdfs://localhost:8020
2018-07-10 23:32:54,319 [main] INFO org.apache.pig.PigServer - Pig Script ID for the session:
PIG-query5.pig-99e30da6-513f-4fbd-ad2b-599c69efba3b
2018-07-10 23:32:54,320 [main] WARN org.apache.pig.PigServer - ATS is disabled since
yarn.timeline-service.enabled set to false
2018-07-10 23:32:55,671 [main] INFO org.apache.hadoop.conf.Configuration.deprecation -
fs.default.name is deprecated. Instead, use fs.defaultFS

```

2018-07-10 23:32:55,841 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS

2018-07-10 23:32:56,169 [main] INFO org.apache.pig.impl.util.SpillableMemoryManager - Selected heap (Tenured Gen) of size 699072512 to monitor. collectionUsageThreshold = 489350752, usageThreshold = 489350752

2018-07-10 23:32:56,189 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: HASH\_JOIN,FILTER

2018-07-10 23:32:56,279 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS

2018-07-10 23:32:56,294 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.

2018-07-10 23:32:56,388 [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]}

2018-07-10 23:32:56,476 [main] INFO org.apache.pig.newplan.logical.rules.ColumnPruneVisitor - Columns pruned for EmpDetail: \$2, \$3

2018-07-10 23:32:56,480 [main] INFO org.apache.pig.newplan.logical.rules.ColumnPruneVisitor - Columns pruned for ExpenseDetail: \$1

2018-07-10 23:32:56,750 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File concatenation threshold: 100 optimistic? false

2018-07-10 23:32:56,874 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size before optimization: 1

2018-07-10 23:32:56,874 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size after optimization: 1

2018-07-10 23:32:56,962 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS

2018-07-10 23:32:57,043 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032

2018-07-10 23:32:57,570 [main] INFO org.apache.pig.tools.pigstats.mapreduce.MRScriptState - Pig script settings are added to the job

2018-07-10 23:32:57,592 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.reduce.markreset.buffer.percent is deprecated. Instead, use mapreduce.reduce.markreset.buffer.percent

2018-07-10 23:32:57,592 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - mapred.job.reduce.markreset.buffer.percent is not set, set to default 0.3

2018-07-10 23:32:57,592 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.output.compress is deprecated. Instead, use mapreduce.output.fileoutputformat.compress

2018-07-10 23:32:57,597 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Reduce phase detected, estimating # of required reducers.

2018-07-10 23:32:57,608 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Using

reducer estimator:

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.InputSizeReducerEstimator  
2018-07-10 23:32:57,668 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.InputSizeReducerEstimator -  
BytesPerReducer=1000000000 maxReducers=999 totalInputFileSize=333

2018-07-10 23:32:57,669 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Setting  
Parallelism to 1

2018-07-10 23:32:57,669 [main] INFO org.apache.hadoop.conf.Configuration.deprecation -  
mapred.reduce.tasks is deprecated. Instead, use mapreduce.job.reduces

2018-07-10 23:32:57,670 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - This job  
cannot be converted run in-process

2018-07-10 23:32:57,745 [main] INFO org.apache.hadoop.conf.Configuration.deprecation -  
mapred.submit.replication is deprecated. Instead, use mapreduce.client.submit.file.replication

2018-07-10 23:32:58,499 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar  
file:/home/acadgild/install/pig/pig-0.16.0/pig-0.16.0-core-h2.jar to DistributedCache through  
/tmp/temp212121560/tmp1037105833/pig-0.16.0-core-h2.jar

2018-07-10 23:32:58,595 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar  
file:/home/acadgild/install/pig/pig-0.16.0/lib/automaton-1.11-8.jar to DistributedCache through  
/tmp/temp212121560/tmp1434209988/automaton-1.11-8.jar

2018-07-10 23:32:58,680 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar  
file:/home/acadgild/install/pig/pig-0.16.0/lib/antlr-runtime-3.4.jar to DistributedCache through  
/tmp/temp212121560/tmp1665493819/antlr-runtime-3.4.jar

2018-07-10 23:32:58,785 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar  
file:/home/acadgild/install/pig/pig-0.16.0/lib/joda-time-2.9.3.jar to DistributedCache through  
/tmp/temp212121560/tmp-1429908685/joda-time-2.9.3.jar

2018-07-10 23:32:58,855 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Setting up  
single store job

2018-07-10 23:32:58,910 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Key  
[pig.schematuple] is false, will not generate code.

2018-07-10 23:32:58,911 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Starting process  
to move generated code to distributed cacche

2018-07-10 23:32:58,913 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Setting key  
[pig.schematuple.classes] with classes to deserialize []

2018-07-10 23:32:59,401 [main] INFO

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

2018-07-10 23:32:59,403 [main] INFO org.apache.hadoop.conf.Configuration.deprecation -  
mapred.job.tracker.http.address is deprecated. Instead, use mapreduce.jobtracker.http.address

2018-07-10 23:32:59,434 [JobControl] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to  
ResourceManager at localhost/127.0.0.1:8032

2018-07-10 23:32:59,537 [JobControl] INFO org.apache.hadoop.conf.Configuration.deprecation -  
mapred.task.id is deprecated. Instead, use mapreduce.task.attempt.id

2018-07-10 23:32:59,559 [JobControl] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS

2018-07-10 23:33:01,253 [JobControl] WARN org.apache.hadoop.mapreduce.JobResourceUploader - No job jar file set. User classes may not be found. See Job or Job#setJar(String).

2018-07-10 23:33:01,406 [JobControl] INFO org.apache.pig.builtin.PigStorage - Using PigTextInputFormat

2018-07-10 23:33:01,484 [JobControl] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1

2018-07-10 23:33:01,484 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1

2018-07-10 23:33:01,654 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths (combined) to process : 1

2018-07-10 23:33:01,664 [JobControl] INFO org.apache.pig.builtin.PigStorage - Using PigTextInputFormat

2018-07-10 23:33:01,672 [JobControl] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1

2018-07-10 23:33:01,672 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1

2018-07-10 23:33:01,689 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths (combined) to process : 1

2018-07-10 23:33:02,017 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter - number of splits:2

2018-07-10 23:33:02,283 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter - Submitting tokens for job: job\_1531229668339\_0078

2018-07-10 23:33:02,587 [JobControl] INFO org.apache.hadoop.mapred.YARNRunner - Job jar is not present. Not adding any jar to the list of resources.

2018-07-10 23:33:02,889 [JobControl] INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl - Submitted application application\_1531229668339\_0078

2018-07-10 23:33:03,129 [JobControl] INFO org.apache.hadoop.mapreduce.Job - The url to track the job: http://localhost:8088/proxy/application\_1531229668339\_0078/

2018-07-10 23:33:03,131 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - HadoopJobId: job\_1531229668339\_0078

2018-07-10 23:33:03,131 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Processing aliases EmpDetail,ExpenseDetail,Sol5,nonCommonEmp,uncommon1

2018-07-10 23:33:03,131 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - detailed locations: M: ExpenseDetail[2,16],ExpenseDetail[-1,-1],nonCommonEmp[3,15],EmpDetail[1,12],EmpDetail[-1,-1],nonCommonEmp[3,15] C: R: nonCommonEmp[3,15],uncommon1[4,12],Sol5[5,7]

2018-07-10 23:33:03,189 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 0% complete

2018-07-10 23:33:03,189 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Running jobs are [job\_1531229668339\_0078]

```

2018-07-10 23:33:40,495 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 50%
complete
2018-07-10 23:33:40,501 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Running
jobs are [job_1531229668339_0078]
2018-07-10 23:33:52,537 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Running
jobs are [job_1531229668339_0078]
2018-07-10 23:33:53,553 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to
ResourceManager at localhost/127.0.0.1:8032
2018-07-10 23:33:53,573 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate -
Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history
server
2018-07-10 23:33:53,972 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to
ResourceManager at localhost/127.0.0.1:8032
2018-07-10 23:33:53,983 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate -
Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history
server
2018-07-10 23:33:54,104 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to
ResourceManager at localhost/127.0.0.1:8032
2018-07-10 23:33:54,111 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate -
Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history
server
2018-07-10 23:33:54,214 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 100%
complete
2018-07-10 23:33:54,225 [main] INFO org.apache.pig.tools.pigstats.mapreduce.SimplePigStats -
Script Statistics:

```

HadoopVersion	PigVersion	UserId	StartedAt	FinishedAt	Features
2.6.5	0.16.0	acadgild	2018-07-10 23:32:57	2018-07-10 23:33:54	HASH_JOIN,FILTER

Success!

Job Stats (time in seconds):

JobId	Maps	Reduces	MaxMapTime	MinMapTime	AvgMapTime	MedianMapTime	MaxReduceTime	MinReduceTime	AvgReduceTime	MedianReductime	Alias	Feature
job_1531229668339_0078	2	1	16	16	16	16	9	9	9	9		
EmpDetail,ExpenseDetail,Sol5,nonCommonEmp,uncommon1												
HASH_JOIN												
hdfs://localhost:8020/tmp/temp212121560/tmp-581297406,												

Input(s):

Successfully read 9 records from: "/user/acadgild/hadoopdata/pig/input/employee\_expenses.txt"  
 Successfully read 14 records from: "/user/acadgild/hadoopdata/pig/input/employee\_details.txt"

Output(s):

Successfully stored 8 records (118 bytes) in:

"hdfs://localhost:8020/tmp/temp212121560/tmp-581297406"

Counters:

Total records written : 8

Total bytes written : 118

Spillable Memory Manager spill count : 0

Total bags proactively spilled: 0

Total records proactively spilled: 0

Job DAG:

job\_1531229668339\_0078

2018-07-10 23:33:54,233 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032

2018-07-10 23:33:54,242 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

2018-07-10 23:33:54,314 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032

2018-07-10 23:33:54,319 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

2018-07-10 23:33:54,387 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032

2018-07-10 23:33:54,395 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

2018-07-10 23:33:54,471 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!

2018-07-10 23:33:54,482 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS

2018-07-10 23:33:54,486 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.

2018-07-10 23:33:54,518 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1

2018-07-10 23:33:54,519 [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)**

2018-07-10 23:33:54,701 [main] INFO org.apache.pig.Main - Pig script completed in 1 minute, 3 seconds and 387 milliseconds (63387 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/>

#### **CSV files loaded to HDFS from terminal :**

```
[acadgild@localhost ~]$ hdfs dfs -put /home/acadgild/Desktop/airports.csv
/user/acadgild/hadoopdata/pig/input
18/07/10 23:16:21 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your
platform... using builtin-java classes where applicable
[acadgild@localhost ~]$ hdfs dfs -put /home/acadgild/Desktop/DelayedFlights.csv
/user/acadgild/hadoopdata/pig/input18/07/10 23:17:14 WARN util.NativeCodeLoader: Unable to load
native-hadoop library for your platform... using builtin-java classes where applicable
[acadgild@localhost ~]$ hdfs dfs -ls /user/acadgild/hadoopdata/pig/input
18/07/10 23:17:54 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your
platform... using builtin-java classes where applicable
Found 5 items
-rw-r--r--  1 acadgild supergroup  247963212 2018-07-10 23:17
/user/acadgild/hadoopdata/pig/input/DelayedFlights.csv
-rw-r--r--  1 acadgild supergroup    244438 2018-07-10 23:16
/user/acadgild/hadoopdata/pig/input/airports.csv
-rw-r--r--  1 acadgild supergroup     261 2018-07-06 16:56
/user/acadgild/hadoopdata/pig/input/employee_details.txt
-rw-r--r--  1 acadgild supergroup     72 2018-07-06 16:57
/user/acadgild/hadoopdata/pig/input/employee_expenses.txt
-rw-r--r--  1 acadgild supergroup     76 2018-07-06 17:27
/user/acadgild/hadoopdata/pig/input/sample.txt
[acadgild@localhost ~]$
```

#### **Problem Statement 1**

Find out the top 5 most visited destinations.

REGISTER '/home/acadgild/Desktop/piggybank.jar';

A = load '/user/acadgild/hadoopdata/pig/input/DelayedFlights.csv' USING

```
org.apache.pig.piggybank.storage.CSVExcelStorage(',', 'NO_MULTILINE', 'UNIX', 'SKIP_INPUT_HEADER');

B = foreach A generate (int)$1 as year, (int)$10 as flight_num, (chararray)$17 as origin, (chararray)$18 as dest;

C = filter B by dest is not null;

D = group C by dest;

E = foreach D generate group, COUNT(C.dest);

F = order E by $1 DESC;

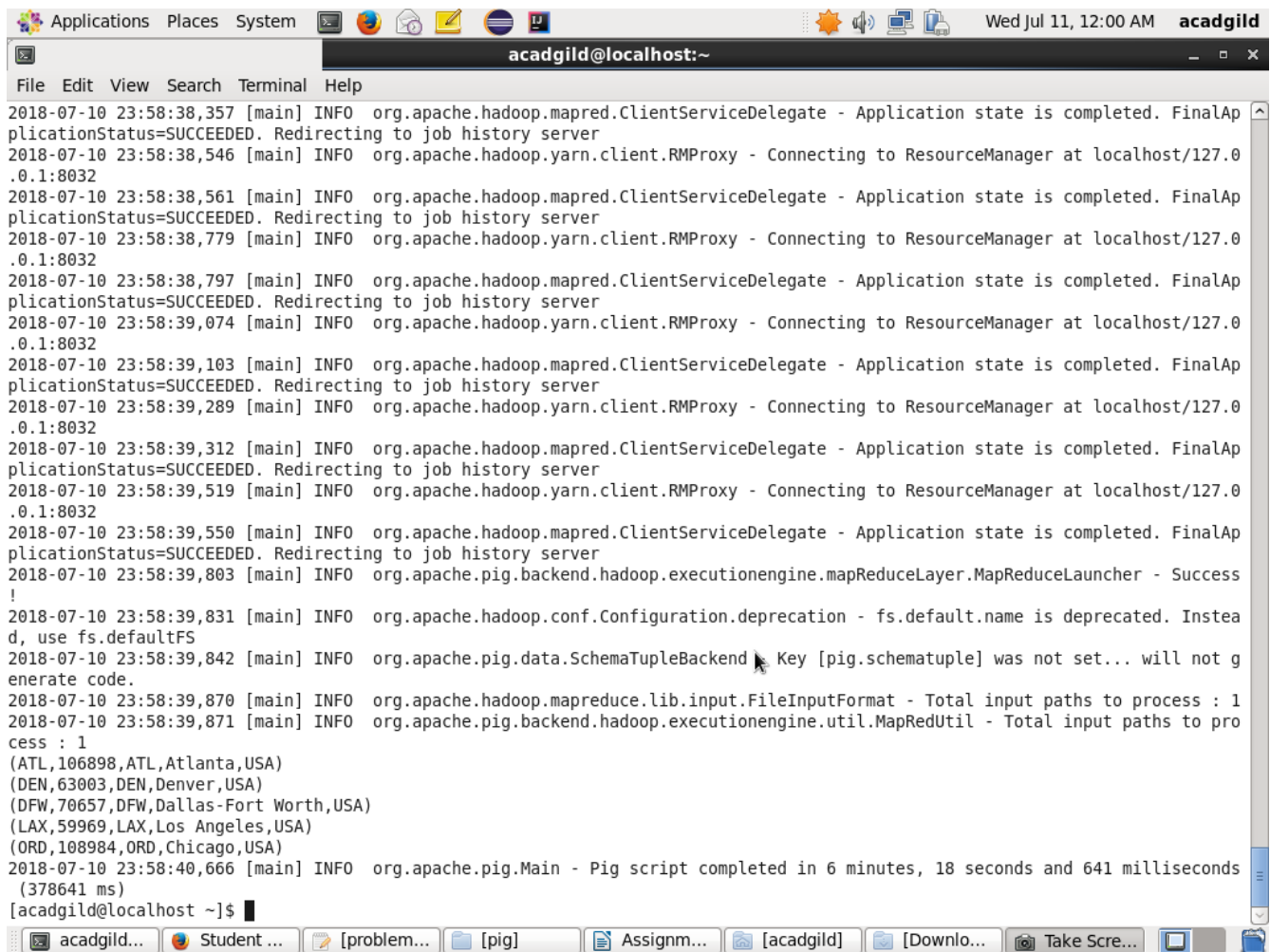
Result = LIMIT F 5;

A1 = load '/user/acadgild/hadoopdata/pig/input/airports.csv' USING
org.apache.pig.piggybank.storage.CSVExcelStorage(',', 'NO_MULTILINE', 'UNIX', 'SKIP_INPUT_HEADER');

A2 = foreach A1 generate (chararray)$0 as dest, (chararray)$2 as city, (chararray)$4 as country;

joined_table = join Result by $0, A2 by dest;

dump joined_table;
```

A screenshot of a Linux terminal window titled 'acadgild@localhost:~'. The window shows a series of log messages from Hadoop and Pig. The logs indicate that several Hadoop MapReduce applications have completed successfully and are redirecting to the job history server. There are also messages from the Pig backend regarding configuration deprecation and a warning about a missing key. The terminal ends with the prompt 'acadgild@localhost ~]\$'. The desktop environment is visible at the top and bottom of the window, showing various application icons and a taskbar.

```
2018-07-10 23:58:38,357 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalAp
plicationStatus=SUCCEEDED. Redirecting to job history server
2018-07-10 23:58:38,546 [main] INFO org.apache.hadoop.yarn.client.RMPProxy - Connecting to ResourceManager at localhost/127.0
.0.1:8032
2018-07-10 23:58:38,561 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalAp
plicationStatus=SUCCEEDED. Redirecting to job history server
2018-07-10 23:58:38,779 [main] INFO org.apache.hadoop.yarn.client.RMPProxy - Connecting to ResourceManager at localhost/127.0
.0.1:8032
2018-07-10 23:58:38,797 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalAp
plicationStatus=SUCCEEDED. Redirecting to job history server
2018-07-10 23:58:39,074 [main] INFO org.apache.hadoop.yarn.client.RMPProxy - Connecting to ResourceManager at localhost/127.0
.0.1:8032
2018-07-10 23:58:39,103 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalAp
plicationStatus=SUCCEEDED. Redirecting to job history server
2018-07-10 23:58:39,289 [main] INFO org.apache.hadoop.yarn.client.RMPProxy - Connecting to ResourceManager at localhost/127.0
.0.1:8032
2018-07-10 23:58:39,312 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalAp
plicationStatus=SUCCEEDED. Redirecting to job history server
2018-07-10 23:58:39,519 [main] INFO org.apache.hadoop.yarn.client.RMPProxy - Connecting to ResourceManager at localhost/127.0
.0.1:8032
2018-07-10 23:58:39,550 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalAp
plicationStatus=SUCCEEDED. Redirecting to job history server
2018-07-10 23:58:39,803 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success
!
2018-07-10 23:58:39,831 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instea
d, use fs.defaultFS
2018-07-10 23:58:39,842 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not g
enerate code.
2018-07-10 23:58:39,870 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2018-07-10 23:58:39,871 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to pro
cess : 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)
2018-07-10 23:58:40,666 [main] INFO org.apache.pig.Main - Pig script completed in 6 minutes, 18 seconds and 641 milliseconds
(378641 ms)
[acadgild@localhost ~]$
```

## Problem Statement 2

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

REGISTER '/home/acadgild/Desktop/piggybank.jar';

A = load '/user/acadgild/hadoopdata/pig/input/DelayedFlights.csv' USING

```
org.apache.pig.piggybank.storage.CSVExcelStorage(',', 'NO_MULTILINE', 'UNIX', 'SKIP_INPUT_HEADER');
```

```
B = foreach A generate (int)$2 as month, (int)$10 as flight_num, (int)$22 as cancelled, (chararray)$23 as cancel_code;
```

```
C = filter B by cancelled == 1 AND cancel_code == 'B';
```

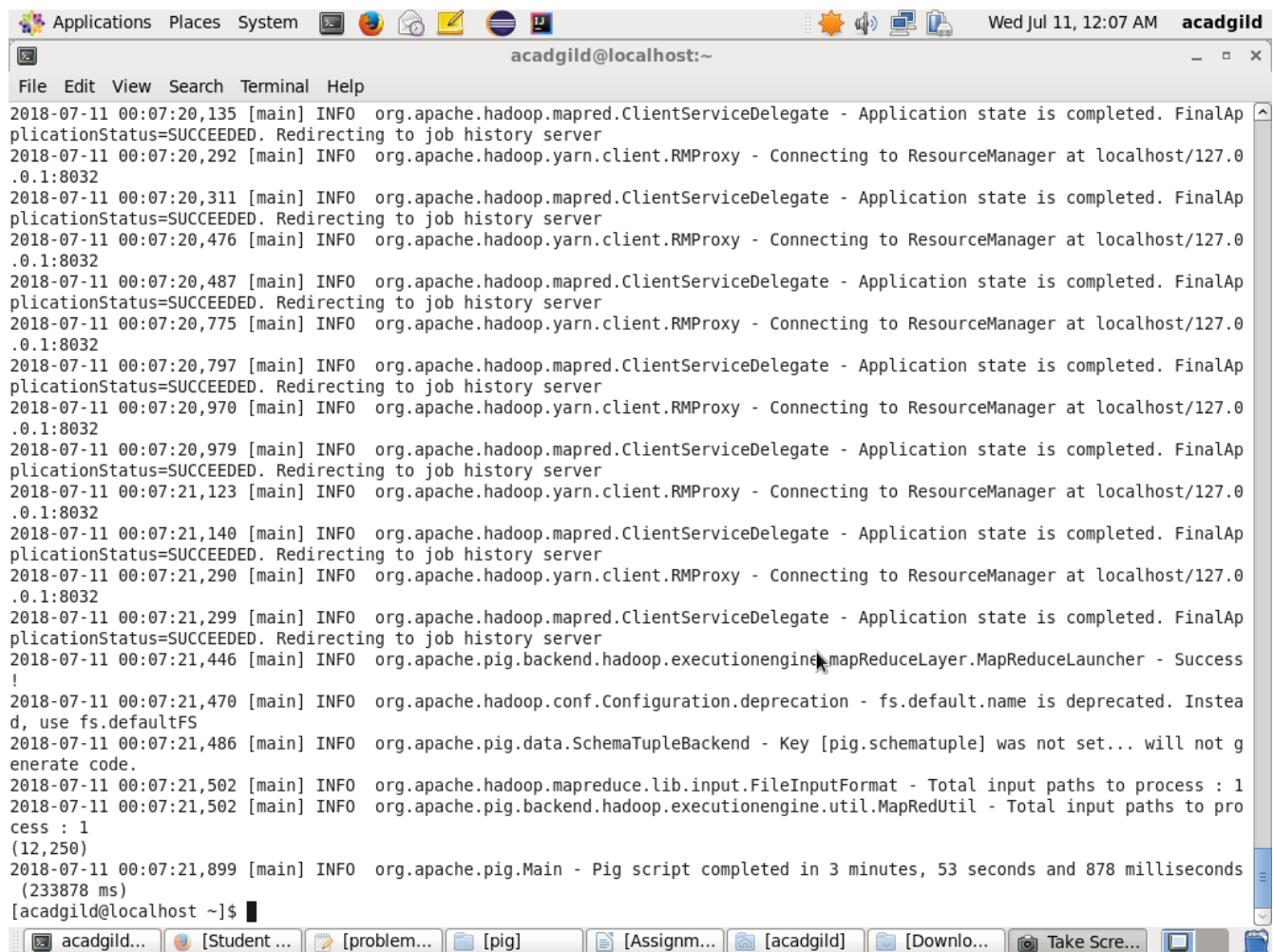
```
D = group C by month;
```

```
E = foreach D generate group, COUNT(C.cancelled);
```

```
F = order E by $1 DESC;
```

```
Result = limit F 1;
```

```
dump Result;
```



The screenshot shows a terminal window titled 'acadgild@localhost:~' with a menu bar (File, Edit, View, Search, Terminal, Help). The terminal displays a series of log messages from Hadoop and Pig. The logs indicate that several MapReduce applications completed successfully and were redirected to the job history server. There are also warnings about deprecated configuration options and a missing key in the Pig schema. The final log message states: '2018-07-11 00:07:21,899 [main] INFO org.apache.pig.Main - Pig script completed in 3 minutes, 53 seconds and 878 milliseconds (233878 ms)'. The prompt '[acadgild@localhost ~]\$' is visible at the bottom.

```
2018-07-11 00:07:20,135 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalAp
plicationStatus=SUCCEEDED. Redirecting to job history server
2018-07-11 00:07:20,292 [main] INFO org.apache.hadoop.yarn.client.RMPProxy - Connecting to ResourceManager at localhost/127.0
.0.1:8032
2018-07-11 00:07:20,311 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalAp
plicationStatus=SUCCEEDED. Redirecting to job history server
2018-07-11 00:07:20,476 [main] INFO org.apache.hadoop.yarn.client.RMPProxy - Connecting to ResourceManager at localhost/127.0
.0.1:8032
2018-07-11 00:07:20,487 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalAp
plicationStatus=SUCCEEDED. Redirecting to job history server
2018-07-11 00:07:20,775 [main] INFO org.apache.hadoop.yarn.client.RMPProxy - Connecting to ResourceManager at localhost/127.0
.0.1:8032
2018-07-11 00:07:20,797 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalAp
plicationStatus=SUCCEEDED. Redirecting to job history server
2018-07-11 00:07:20,970 [main] INFO org.apache.hadoop.yarn.client.RMPProxy - Connecting to ResourceManager at localhost/127.0
.0.1:8032
2018-07-11 00:07:20,979 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalAp
plicationStatus=SUCCEEDED. Redirecting to job history server
2018-07-11 00:07:21,123 [main] INFO org.apache.hadoop.yarn.client.RMPProxy - Connecting to ResourceManager at localhost/127.0
.0.1:8032
2018-07-11 00:07:21,140 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalAp
plicationStatus=SUCCEEDED. Redirecting to job history server
2018-07-11 00:07:21,290 [main] INFO org.apache.hadoop.yarn.client.RMPProxy - Connecting to ResourceManager at localhost/127.0
.0.1:8032
2018-07-11 00:07:21,299 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalAp
plicationStatus=SUCCEEDED. Redirecting to job history server
2018-07-11 00:07:21,446 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success
!
2018-07-11 00:07:21,470 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instea
d, use fs.defaultFS
2018-07-11 00:07:21,486 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not g
enerate code.
2018-07-11 00:07:21,502 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2018-07-11 00:07:21,502 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to pro
cess : 1
(12,250)
2018-07-11 00:07:21,899 [main] INFO org.apache.pig.Main - Pig script completed in 3 minutes, 53 seconds and 878 milliseconds
(233878 ms)
[acadgild@localhost ~]$
```

### Problem Statement 3

Top ten origins with the highest AVG departure delay

```
REGISTER '/home/acadgild/Desktop/piggybank.jar';
```

```
A = load '/user/acadgild/hadoopdata/pig/input/DelayedFlights.csv' USING  
org.apache.pig.piggybank.storage.CSVExcelStorage(',', 'NO_MULTILINE', 'UNIX', 'SKIP_INPUT_HEADER');
```

```
B1 = foreach A generate (int)$16 as dep_delay, (chararray)$17 as origin;
```

```
C1 = filter B1 by (dep_delay is not null) AND (origin is not null);
```

```
D1 = group C1 by origin;
```

```
E1 = foreach D1 generate group, AVG(C1.dep_delay);
```

```
Result = order E1 by $1 DESC;
```

```
Top_ten = limit Result 10;
```

```
Lookup = load '/user/acadgild/hadoopdata/pig/input/airports.csv' USING  
org.apache.pig.piggybank.storage.CSVExcelStorage(',', 'NO_MULTILINE', 'UNIX', 'SKIP_INPUT_HEADER');
```

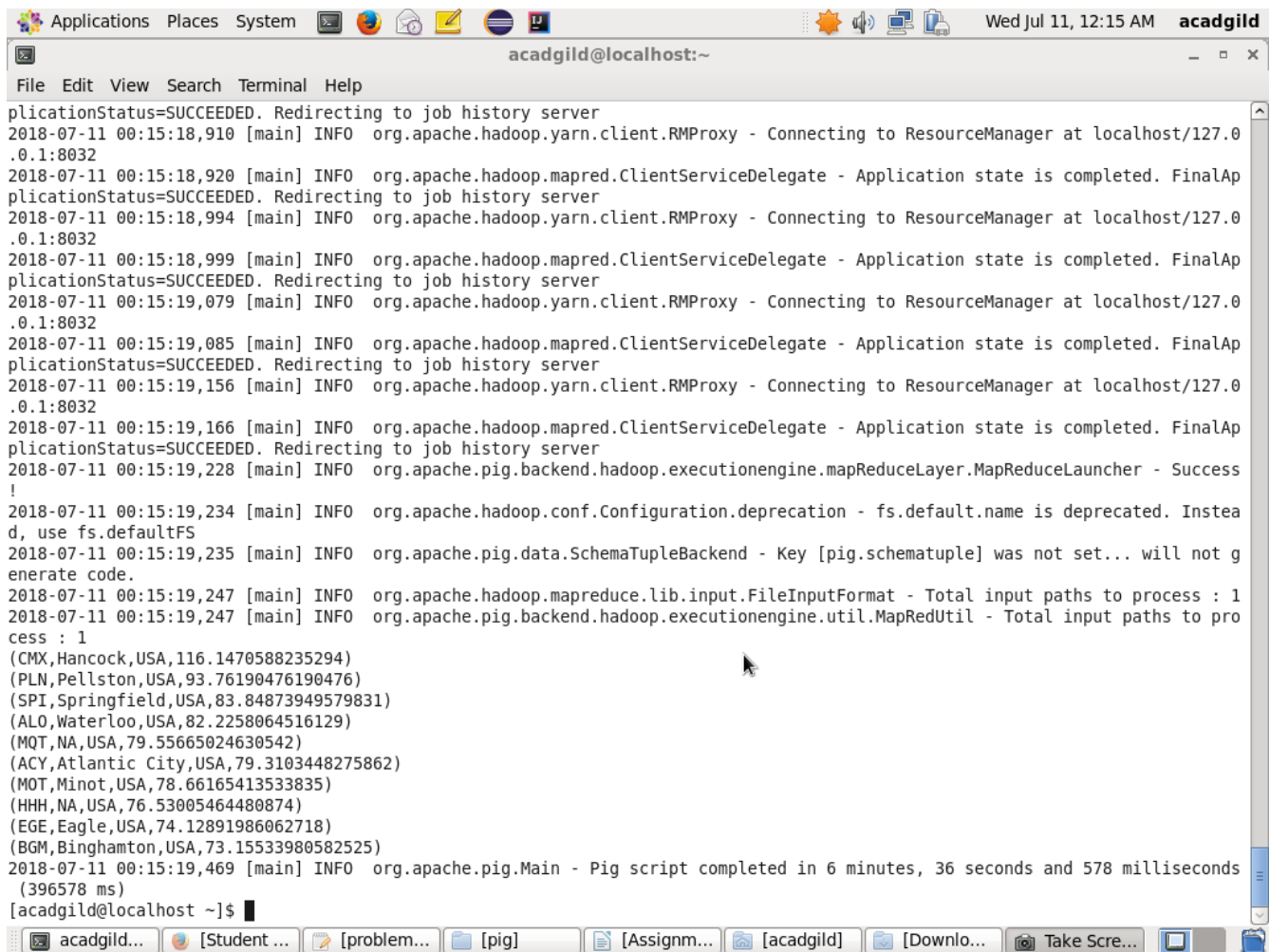
```
Lookup1 = foreach Lookup generate (chararray)$0 as origin, (chararray)$2 as city, (chararray)$4 as  
country;
```

```
Joined = join Lookup1 by origin, Top_ten by $0;
```

```
Final = foreach Joined generate $0,$1,$2,$4;
```

```
Final_Result = ORDER Final by $3 DESC;
```

```
dump Final_Result;
```



```
plicationStatus=SUCCEEDED. Redirecting to job history server
2018-07-11 00:15:18,910 [main] INFO org.apache.hadoop.yarn.client.RMPProxy - Connecting to ResourceManager at localhost/127.0
.0.1:8032
2018-07-11 00:15:18,920 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalAp
plicationStatus=SUCCEEDED. Redirecting to job history server
2018-07-11 00:15:18,994 [main] INFO org.apache.hadoop.yarn.client.RMPProxy - Connecting to ResourceManager at localhost/127.0
.0.1:8032
2018-07-11 00:15:18,999 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalAp
plicationStatus=SUCCEEDED. Redirecting to job history server
2018-07-11 00:15:19,079 [main] INFO org.apache.hadoop.yarn.client.RMPProxy - Connecting to ResourceManager at localhost/127.0
.0.1:8032
2018-07-11 00:15:19,085 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalAp
plicationStatus=SUCCEEDED. Redirecting to job history server
2018-07-11 00:15:19,156 [main] INFO org.apache.hadoop.yarn.client.RMPProxy - Connecting to ResourceManager at localhost/127.0
.0.1:8032
2018-07-11 00:15:19,166 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalAp
plicationStatus=SUCCEEDED. Redirecting to job history server
2018-07-11 00:15:19,228 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success
!
2018-07-11 00:15:19,234 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instea
d, use fs.defaultFS
2018-07-11 00:15:19,235 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not g
enerate code.
2018-07-11 00:15:19,247 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2018-07-11 00:15:19,247 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to pro
cess : 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,76.53005464480874)
(EGE,Eagle,USA,74.12891986062718)
(BGM,Binghamton,USA,73.15533980582525)
2018-07-11 00:15:19,469 [main] INFO org.apache.pig.Main - Pig script completed in 6 minutes, 36 seconds and 578 milliseconds
(396578 ms)
[acadgild@localhost ~]$
```

## Problem Statement 4

Which route (origin & destination) has seen the maximum diversion?

```
REGISTER '/home/acadgild/Desktop/piggybank.jar';
```

```
A = load '/user/acadgild/hadoopdata/pig/input/DelayedFlights.csv' USING
org.apache.pig.piggybank.storage.CSVExcelStorage(',', 'NO_MULTILINE', 'UNIX', 'SKIP_INPUT_HE
```

ADER');

B = FOREACH A GENERATE (chararray)\$17 as origin, (chararray)\$18 as dest, (int)\$24 as diversion;

C = FILTER B BY (origin is not null) AND (dest is not null) AND (diversion == 1);

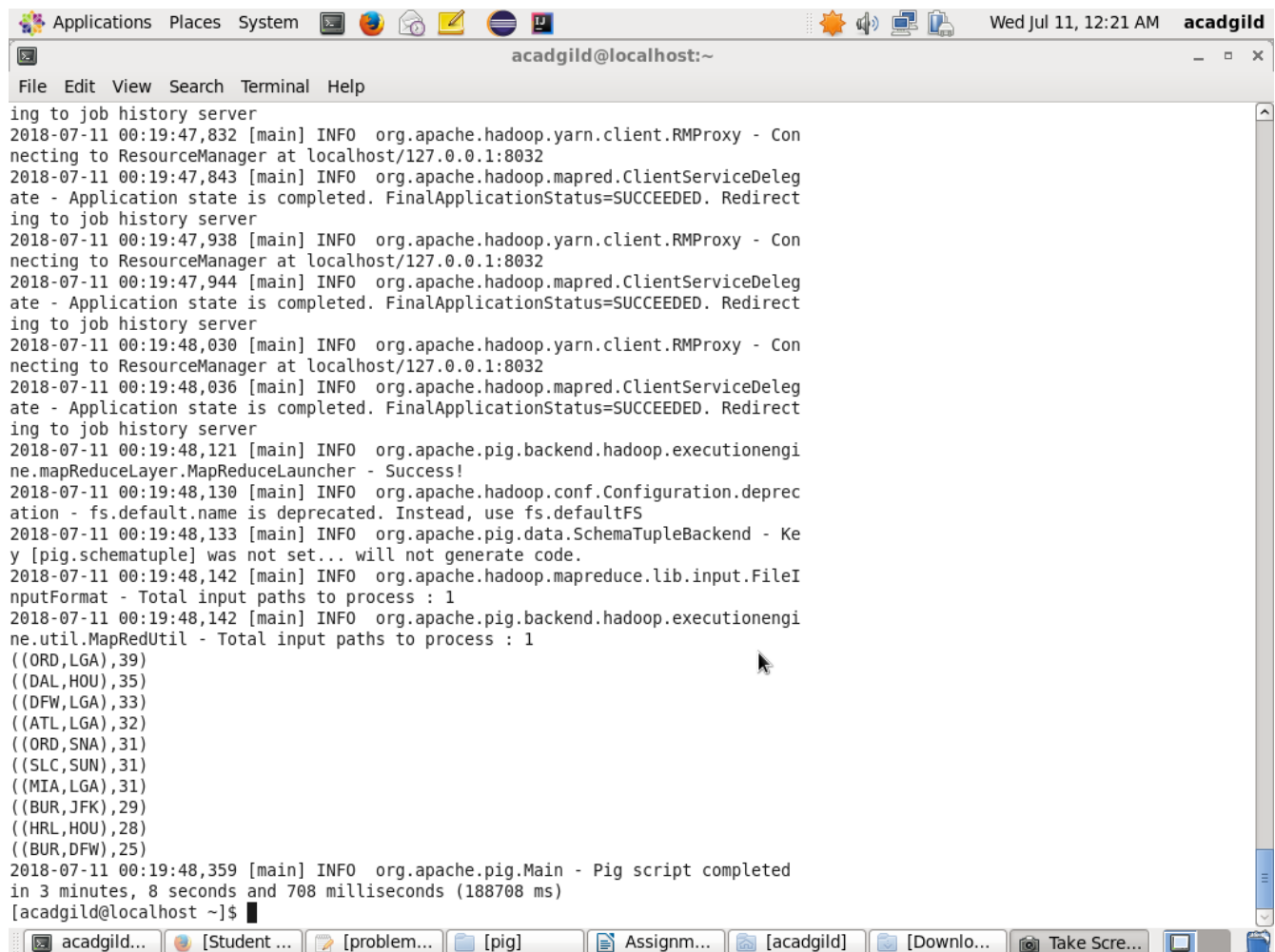
D = GROUP C by (origin,dest);

E = FOREACH D generate group, COUNT(C.diversion);

F = ORDER E BY \$1 DESC;

Result = limit F 10;

dump Result;



The screenshot shows a terminal window titled 'acadgild@localhost:~' with a menu bar (File, Edit, View, Search, Terminal, Help). The terminal output displays logs from Hadoop and Pig. It includes messages about connecting to the ResourceManager, application completion, and Pig script execution. The final output of the Pig script is a list of airport pairs and their diversion counts, sorted in descending order. The logs end with the prompt '[acadgild@localhost ~]\$'.

```
ing to job history server
2018-07-11 00:19:47,832 [main] INFO org.apache.hadoop.yarn.client.RMPProxy - Con
necting to ResourceManager at localhost/127.0.0.1:8032
2018-07-11 00:19:47,843 [main] INFO org.apache.hadoop.mapred.ClientServiceDeleg
ate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirect
ing to job history server
2018-07-11 00:19:47,938 [main] INFO org.apache.hadoop.yarn.client.RMPProxy - Con
necting to ResourceManager at localhost/127.0.0.1:8032
2018-07-11 00:19:47,944 [main] INFO org.apache.hadoop.mapred.ClientServiceDeleg
ate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirect
ing to job history server
2018-07-11 00:19:48,030 [main] INFO org.apache.hadoop.yarn.client.RMPProxy - Con
necting to ResourceManager at localhost/127.0.0.1:8032
2018-07-11 00:19:48,036 [main] INFO org.apache.hadoop.mapred.ClientServiceDeleg
ate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirect
ing to job history server
2018-07-11 00:19:48,121 [main] INFO org.apache.pig.backend.hadoop.executionengi
ne.mapReduceLayer.MapReduceLauncher - Success!
2018-07-11 00:19:48,130 [main] INFO org.apache.hadoop.conf.Configuration.deprec
ation - fs.default.name is deprecated. Instead, use fs.defaultFS
2018-07-11 00:19:48,133 [main] INFO org.apache.pig.data.SchemaTupleBackend - Ke
y [pig.schematuple] was not set... will not generate code.
2018-07-11 00:19:48,142 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileI
nputFormat - Total input paths to process : 1
2018-07-11 00:19:48,142 [main] INFO org.apache.pig.backend.hadoop.executionengi
ne.util.MapRedUtil - 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)
2018-07-11 00:19:48,359 [main] INFO org.apache.pig.Main - Pig script completed
in 3 minutes, 8 seconds and 708 milliseconds (188708 ms)
[acadgild@localhost ~]$
```