AIM: Implementation of PIG commands.

OUTPUT:

1. Run PIG –

```
[cloudera@quickstart ~]$ pig
log4j:WARN No appenders could be found for logger (org.apache.hadoop.util.Shell).
log4j:WARN Please initialize the log4j system properly.
log4j:WARN See http://logging.apache.org/log4j/1.2/faq.html#noconfig for more info.
2020-11-25 03:57:25,756 [main] INFO  org.apache.pig.Main - Apache Pig version 0.12.0-cdh5.12.0 (rexported) compiled Jun 29 2017, 04:34:31
2020-11-25 03:57:25,757 [main] INFO  org.apache.pig.Main - Logging error messages to: /home/cloudera/pig_1606305445677.log
2020-11-25 03:57:25,825 [main] INFO  org.apache.pig.impl.util.Utils - Default bootup file /home/cloudera/.pigbootup not found
2020-11-25 03:57:28,595 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2020-11-25 03:57:28,595 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2020-11-25 03:57:28,595 [main] INFO  org.apache.pig.backend.hadoop.executionengine.HExecutionEngine - Connecting to hadoop file system at: hdfs://quickstart.cloudera:80
20
2020-11-25 03:57:34,057 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2020-11-25 03:57:34,057 [main] INFO  org.apache.pig.backend.hadoop.executionengine.HExecutionEngine - Connecting to map-reduce job tracker at: localhost:8021
2020-11-25 03:57:34,059 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2020-11-25 03:57:34,339 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2020-11-25 03:57:34,345 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2020-11-25 03:57:34,695 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2020-11-25 03:57:34,705 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2020-11-25 03:57:35,092 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2020-11-25 03:57:35,105 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2020-11-25 03:57:35,459 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2020-11-25 03:57:35,460 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2020-11-25 03:57:35,870 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2020-11-25 03:57:35,873 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2020-11-25 03:57:36,205 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2020-11-25 03:57:36,214 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2020-11-25 03:57:36,562 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2020-11-25 03:57:36,568 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2020-11-25 03:57:36,919 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2020-11-25 03:57:36,926 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
grunt>
```

2. fs –ls –

```
grunt> mkdir trainP
grunt> fs -ls
Found 2 items
drwxr-xr-x   - cloudera cloudera          0 2020-11-25 04:08 t2
drwxr-xr-x   - cloudera cloudera          0 2020-11-25 04:13 trainP
```

3. copyFromLocal –

```
grunt> copyFromLocal /home/cloudera/trainingPracs/demoP.txt /trainP
grunt> cat /user/cloudera/trainP/demoP.txt
1,abc,2000,Admin
2,def,4500,Tech
3,ghi,5600,PR
4,jkl,8000,Tech
5,mno,2300,Admin
grunt>
```

4. Dump –

```
grunt> student = load '/user/cloudera/trainP/students.txt' USING PigStorage(',') as (rollno:int,name:chararray,percentage:float,sex:chararray);
2020-11-25 04:26:15,348 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2020-11-25 04:26:15,348 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
grunt>
```

```
grunt> dump student;
2020-11-25 04:26:36,978 [main] INFO  org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: UNKNOWN
2020-11-25 04:26:37,138 [main] INFO  org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - {RULES ENABLED=[AddForEach, ColumnMapKeyPrune, DuplicateForEachColu
mnRewrite, GroupByConstParallelSetter, ImplicitSplitInserter, LimitOptimizer, LoadTypeCastInserter, MergeFilter, MergeForEach, NewPartitionFilterOptimizer, PushDownForE
achFlatten, PushUpFilter, SplitFilter, StreamTypeCastInserter], RULES_DISABLED=[FilterLogicExpressionSimplifier, PartitionFilterOptimizer]}
2020-11-25 04:26:37,553 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File concatenation threshold: 100 optimistic? false
2020-11-25 04:26:37,648 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size before optimization: 1
2020-11-25 04:26:37,648 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size after optimization: 1
2020-11-25 04:26:37,982 [main] INFO  org.apache.pig.tools.pigstats.ScriptState - Pig script settings are added to the job
2020-11-25 04:26:38,929 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapred.job.reduce.markreset.buffer.percent is deprecated. Instead, use mapreduce
.reduce.markreset.buffer.percent
2020-11-25 04:26:38,929 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - mapred.job.reduce.markreset.buffer.percent is not
 set, set to default 0.3
2020-11-25 04:26:38,929 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapred.output.compress is deprecated. Instead, use mapreduce.output.fileoutputfo
rmat.compress
2020-11-25 04:26:56,356 [JobControl] INFO  org.apache.hadoop.mapreduce.JobSubmitter - number of splits:1
2020-11-25 04:26:57,414 [JobControl] INFO  org.apache.hadoop.mapreduce.JobSubmitter - Submitting tokens for job: job_1606301819383_0006
2020-11-25 04:26:58,439 [JobControl] INFO  org.apache.hadoop.yarn.client.api.impl.YarnClientImpl - Submitted application application_1606301819383_0006
2020-11-25 04:26:58,783 [JobControl] INFO  org.apache.hadoop.mapreduce.Job - The url to track the job: http://quickstart.cloudera:8088/proxy/application_1606301819383_0
006/
2020-11-25 04:26:58,783 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - HadoopJobId: job_1606301819383_0006
2020-11-25 04:26:58,783 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Processing aliases student
2020-11-25 04:26:58,784 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - detailed locations: M: student[1,10],student[-1,-1
] C:  R:
2020-11-25 04:26:58,784 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - More information at: http://localhost:50030/jobdet
ails.jsp?jobid=job_1606301819383_0006
2020-11-25 04:26:58,907 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 0% complete
2020-11-25 04:27:52,501 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 50% complete
2020-11-25 04:27:54,733 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapred.reduce.tasks is deprecated. Instead, use mapreduce.job.reduces
2020-11-25 04:27:54,918 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 100% complete
2020-11-25 04:27:54,923 [main] INFO  org.apache.pig.tools.pigstats.SimplePigStats - Script Statistics:

HadoopVersion  PigVersion      UserId  StartedAt       FinishedAt      Features
2.6.0-cdh5.12.0 0.12.0-cdh5.12.0        cloudera        2020-11-25 04:26:38     2020-11-25 04:27:54     UNKNOWN

Success!

Job Stats (time in seconds):
JobId   Maps    Reduces MaxMapTime      MinMapTime      AvgMapTime      MedianMapTime   MaxReduceTime   MinReduceTime   AvgReduceTime   MedianReducetime        Alias  F
eature  Outputs
job_1606301819383_0006  1       0       19      19      19      19      n/a     n/a     n/a     n/a     student MAP_ONLY        hdfs://quickstart.cloudera:8020/tmp/temp
-1340503213/tmp553182692,

Input(s):
Successfully read 9 records (509 bytes) from: "/user/cloudera/trainP/students.txt"

Output(s):
Successfully stored 9 records (216 bytes) in: "hdfs://quickstart.cloudera:8020/tmp/temp-1340503213/tmp553182692"

Counters:
Total records written : 9
Total bytes written : 216
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_1606301819383_0006


2020-11-25 04:27:55,460 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2020-11-25 04:27:55,473 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2020-11-25 04:27:55,475 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2020-11-25 04:27:55,479 [main] INFO  org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.
2020-11-25 04:27:55,531 [main] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2020-11-25 04:27:55,531 [main] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(1,sushil,68.2,F)
(2,deepak,70.33,F)
(3,siddhesh,80.21,F)
(4,vikas,78.3,F)
(5,sumit,62.0,F)
(6,ankita,68.0,F)
(7,supriya,72.21,F)
(8,neha,91.22,F)
(9,mayuresh,85.66,F)
grunt>
```

5. Projection:-

```
grunt> studentname = foreach student generate name;
grunt> dump studentname;
2020-11-25 04:31:16,512 [main] INFO  org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: UNKNOWN
2020-11-25 04:31:16,515 [main] INFO  org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - {RULES ENABLED=[AddForEach, ColumnMapKeyPrune, DuplicateForEachColu
mnRewrite, GroupByConstParallelSetter, ImplicitSplitInserter, LimitOptimizer, LoadTypeCastInserter, MergeFilter, MergeForEach, NewPartitionFilterOptimizer, PushDownForE
achFlatten, PushUpFilter, SplitFilter, StreamTypeCastInserter], RULES_DISABLED=[FilterLogicExpressionSimplifier, PartitionFilterOptimizer]}
2020-11-25 04:31:16,531 [main] INFO  org.apache.pig.newplan.logical.rules.ColumnPruneVisitor - Columns pruned for student: $0, $2, $3
2020-11-25 04:31:16,555 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File concatenation threshold: 100 optimistic? false
2020-11-25 04:31:16,558 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size before optimization: 1
2020-11-25 04:31:16,558 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size after optimization: 1
2020-11-25 04:31:16,645 [main] INFO  org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2020-11-25 04:31:16,652 [main] INFO  org.apache.pig.tools.pigstats.ScriptState - Pig script settings are added to the job
2020-11-25 04:31:16,682 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - mapred.job.reduce.markreset.buffer.percent is not
 set, set to default 0.3
```

```
2020-11-25 04:31:17,551 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - creating jar file Job426848800511787951.jar
2020-11-25 04:31:30,427 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - jar file Job426848800511787951.jar created
2020-11-25 04:31:30,534 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Setting up single store job
2020-11-25 04:31:30,536 [main] INFO  org.apache.pig.data.SchemaTupleFrontend - Key [pig.schematuple] is false, will not generate code.
2020-11-25 04:31:30,536 [main] INFO  org.apache.pig.data.SchemaTupleFrontend - Starting process to move generated code to distributed cache
2020-11-25 04:31:30,537 [main] INFO  org.apache.pig.data.SchemaTupleFrontend - Setting key [pig.schematuple.classes] with classes to deserialize []
2020-11-25 04:31:30,609 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 1 map-reduce job(s) waiting for submission.
2020-11-25 04:31:30,609 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2020-11-25 04:31:30,615 [JobControl] INFO  org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2020-11-25 04:31:30,644 [JobControl] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2020-11-25 04:31:31,258 [JobControl] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2020-11-25 04:31:31,259 [JobControl] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
2020-11-25 04:31:31,265 [JobControl] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths (combined) to process : 1
2020-11-25 04:31:31,363 [DataStreamer for file /tmp/hadoop-yarn/staging/cloudera/.staging/job_1606301819383_0007/job.splitmetainfo] WARN  org.apache.hadoop.hdfs.DFSClie
nt - Caught exception
java.lang.InterruptedException
        at java.lang.Object.wait(Native Method)
        at java.lang.Thread.join(Thread.java:1281)
        at java.lang.Thread.join(Thread.java:1355)
        at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.closeResponder(DFSOutputStream.java:952)
        at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.endBlock(DFSOutputStream.java:690)
        at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.run(DFSOutputStream.java:879)
2020-11-25 04:31:31,363 [JobControl] INFO  org.apache.hadoop.mapreduce.JobSubmitter - number of splits:1
2020-11-25 04:31:31,687 [JobControl] INFO  org.apache.hadoop.mapreduce.JobSubmitter - Submitting tokens for job: job_1606301819383_0007
2020-11-25 04:31:31,916 [JobControl] INFO  org.apache.hadoop.yarn.client.api.impl.YarnClientImpl - Submitted application application_1606301819383_0007
2020-11-25 04:31:31,930 [JobControl] INFO  org.apache.hadoop.mapreduce.Job - The url to track the job: http://quickstart.cloudera:8088/proxy/application_1606301819383_0
007/
2020-11-25 04:31:31,930 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - HadoopJobId: job_1606301819383_0007
2020-11-25 04:31:31,931 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Processing aliases student,studentname
2020-11-25 04:31:31,931 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - detailed locations: M: student[1,10],studentname[-
1,-1] C:  R:
2020-11-25 04:31:31,931 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - More information at: http://localhost:50030/jobdet
ails.jsp?jobid=job_1606301819383_0007
2020-11-25 04:31:32,177 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 0% complete
2020-11-25 04:32:20,027 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 50% complete

2020-11-25 04:32:22,609 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 100% complete
2020-11-25 04:32:22,610 [main] INFO  org.apache.pig.tools.pigstats.SimplePigStats - Script Statistics:

HadoopVersion   PigVersion      UserId  StartedAt       FinishedAt      Features
2.6.0-cdh5.12.0 0.12.0-cdh5.12.0        cloudera        2020-11-25 04:31:16     2020-11-25 04:32:22     UNKNOWN

Success!

Job Stats (time in seconds):
JobId   Maps    Reduces MaxMapTime      MinMapTime      AvgMapTime      MedianMapTime   MaxReduceTime   MinReduceTime   AvgReduceTime   MedianReducetime        Alias   F
eature  Outputs
job_1606301819383_0007 1        0       19      19      19      19      n/a     n/a     n/a     n/a     student,studentname     MAP_ONLY        hdfs://quickstart.cloude
ra:8020/tmp/temp-1340503213/tmp1940330106,

Input(s):
Successfully read 9 records (589 bytes) from: "/user/cloudera/trainP/students.txt"

Output(s):
Successfully stored 9 records (118 bytes) in: "hdfs://quickstart.cloudera:8020/tmp/temp-1340503213/tmp1940330106"

Counters:
Total records written : 9
Total bytes written : 118
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_1606301819383_0007


2020-11-25 04:32:22,853 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2020-11-25 04:32:22,854 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2020-11-25 04:32:22,854 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2020-11-25 04:32:22,860 [main] INFO  org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.
2020-11-25 04:32:22,878 [main] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2020-11-25 04:32:22,878 [main] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(sushil)
(deepak)
(siddhesh)
(vikas)
(sumit)
(ankita)
(supriya)
(neha)
(mayuresh)
grunt> 
```

6. <u>Joins–</u>

```
grunt> cat /user/cloudera/trainP/dept.txt
10,sushil,engineer,50000
10,nikhil,engineer,45000
10,supriya,engineer,50000
10,siddesh,engineer,44000
20,manish,clerk,20000
20,mahesh,clerk,25000
30,minal,scientist,60000
30,krishna,scientist,80000
30,govind,scientist,60000
50,rahul,chemist,80000
40,neha,biochemist,90000
grunt> 
```

```
grunt> dept = load '/user/cloudera/trainP/dept.txt' USING PigStorage(',')as(id:int,name:chararray,deptname:chararray,sal:int);
grunt> dump dept;
2020-11-25 04:37:31,874 [main] INFO  org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: UNKNOWN
2020-11-25 04:37:31,876 [main] INFO  org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - {RULES ENABLED=[AddForEach, ColumnMapKeyPrune, DuplicateForEachColu
mnRewrite, GroupByConstParallelSetter, ImplicitSplitInserter, LimitOptimizer, LoadTypeCastInserter, MergeFilter, MergeForEach, NewPartitionFilterOptimizer, PushDownForE
achFlatten, PushUpFilter, SplitFilter, StreamTypeCastInserter], RULES DISABLED=[FilterLogicExpressionSimplifier, PartitionFilterOptimizer]}
2020-11-25 04:37:31,897 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File concatenation threshold: 100 optimistic? false
2020-11-25 04:37:31,910 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size before optimization: 1
2020-11-25 04:37:31,910 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size after optimization: 1
2020-11-25 04:37:32,037 [main] INFO  org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2020-11-25 04:37:32,041 [main] INFO  org.apache.pig.tools.pigstats.ScriptState - Pig script settings are added to the job
2020-11-25 04:37:32,058 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - mapred.job.reduce.markreset.buffer.percent is not
 set, set to default 0.3
2020-11-25 04:37:32,093 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - creating jar file Job1910742779810010191.jar
2020-11-25 04:37:44,679 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - jar file Job1910742779810010191.jar created
2020-11-25 04:37:44,700 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Setting up single store job
2020-11-25 04:37:44,701 [main] INFO  org.apache.pig.data.SchemaTupleFrontend - Key [pig.schematuple] is false, will not generate code.
2020-11-25 04:37:44,701 [main] INFO  org.apache.pig.data.SchemaTupleFrontend - Starting process to move generated code to distributed cache
2020-11-25 04:37:44,701 [main] INFO  org.apache.pig.data.SchemaTupleFrontend - Setting key [pig.schematuple.classes] with classes to deserialize []
2020-11-25 04:37:44,742 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 1 map-reduce job(s) waiting for submission.
2020-11-25 04:37:44,742 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2020-11-25 04:37:44,746 [JobControl] INFO  org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2020-11-25 04:37:44,796 [JobControl] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2020-11-25 04:37:45,587 [JobControl] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2020-11-25 04:37:45,588 [JobControl] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
2020-11-25 04:37:45,592 [JobControl] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths (combined) to process : 1
2020-11-25 04:37:45,664 [JobControl] INFO  org.apache.hadoop.mapreduce.JobSubmitter - number of splits:1
2020-11-25 04:37:45,830 [JobControl] INFO  org.apache.hadoop.mapreduce.JobSubmitter - Submitting tokens for job: job_1606301819383_0010
2020-11-25 04:37:45,937 [JobControl] INFO  org.apache.hadoop.yarn.client.api.impl.YarnClientImpl - Submitted application application_1606301819383_0010
2020-11-25 04:37:45,943 [JobControl] INFO  org.apache.hadoop.mapreduce.Job - The url to track the job: http://quickstart.cloudera:8088/proxy/application_1606301819383_0
010/
2020-11-25 04:37:45,943 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - HadoopJobId: job_1606301819383_0010
2020-11-25 04:37:45,943 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Processing aliases dept
2020-11-25 04:37:45,943 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - detailed locations: M: dept[5,7],dept[-1,-1] C:  R
:
2020-11-25 04:37:45,943 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - More information at: http://localhost:50030/jobdet
ails.jsp?jobid=job_1606301819383_0010
2020-11-25 04:37:46,042 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 0% complete
2020-11-25 04:38:37,003 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 50% complete
2020-11-25 04:38:41,429 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 100% complete
2020-11-25 04:38:41,429 [main] INFO  org.apache.pig.tools.pigstats.SimplePigStats - Script Statistics:

HadoopVersion  PigVersion  UserId  StartedAt        FinishedAt       Features
2.6.0-cdh5.12.0  0.12.0-cdh5.12.0  cloudera    2020-11-25 04:37:32    2020-11-25 04:38:41    UNKNOWN

Success!

Job Stats (time in seconds):
JobId  Maps  Reduces  MaxMapTime  MinMapTime  AvgMapTime  MedianMapTime  MaxReduceTime  MinReduceTime  AvgReduceTime  MedianReducetime  Alias  F
eature  Outputs
job_1606301819383_0010  1      8      21      21      21      21      n/a      n/a      n/a      n/a      dept  MAP_ONLY      hdfs://quickstart.cloudera:8020/tmp/temp
-1340583213/tmp900694038,

Input(s):
Successfully read 11 records (652 bytes) from: "/user/cloudera/trainP/dept.txt"

Output(s):
Successfully stored 11 records (334 bytes) in: "hdfs://quickstart.cloudera:8020/tmp/temp-1340583213/tmp900694038"

Counters:
Total records written : 11

Total bytes written : 334
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_1606301819383_0010


2020-11-25 04:38:41,671 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2020-11-25 04:38:41,672 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2020-11-25 04:38:41,672 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2020-11-25 04:38:41,673 [main] INFO  org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.
2020-11-25 04:38:41,688 [main] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2020-11-25 04:38:41,688 [main] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(10,sushil,engineer,50000)
(10,nikhil,engineer,45000)
(10,supriya,engineer,50000)
(10,siddesh,engineer,44000)
(20,manish,clerk,20000)
(20,mahesh,clerk,25000)
(30,minal,scientist,60000)
(30,krishna,scientist,80000)
(30,govind,scientist,60000)
(50,rahul,chemist,80000)
(40,neha,biochemist,90000)
grunt> █
```

7. <u>Relational Operrators –</u>
   a. **Cross**: The cross operator is userd to calculate the cross product of two or more relations.

```
grunt> x = cross student,dept;
grunt> dump x;
2020-11-25 04:42:07,381 [main] INFO  org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: CROSS
2020-11-25 04:42:07,382 [main] INFO  org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - {RULES ENABLED=[AddForEach, ColumnMapKeyPrune, DuplicateForEachColu
mnRewrite, GroupByConstParallelSetter, ImplicitSplitInserter, LimitOptimizer, LoadTypeCastInserter, MergeFilter, MergeForEach, NewPartitionFilterOptimizer, PushDownForE
achFlatten, PushUpFilter, SplitFilter, StreamTypeCastInserter], RULES DISABLED=[FilterLogicExpressionSimplifier, PartitionFilterOptimizer]}
2020-11-25 04:42:07,434 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File concatenation threshold: 100 optimistic? false
2020-11-25 04:42:07,472 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler$LastInputStreamingOptimizer - Rewrite: POPackage->POForEach
 to POJoinPackage
2020-11-25 04:42:07,474 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size before optimization: 1
2020-11-25 04:42:07,474 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size after optimization: 1
2020-11-25 04:42:07,566 [main] INFO  org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2020-11-25 04:42:07,571 [main] INFO  org.apache.pig.tools.pigstats.ScriptState - Pig script settings are added to the job
2020-11-25 04:42:07,611 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - mapred.job.reduce.markreset.buffer.percent is not
 set, set to default 0.3
2020-11-25 04:42:07,612 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Reduce phase detected, estimating # of required r
educers.
2020-11-25 04:42:07,615 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Using reducer estimator: org.apache.pig.backend.h
adoop.executionengine.mapReduceLayer.InputSizeReducerEstimator
2020-11-25 04:42:07,624 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.InputSizeReducerEstimator - BytesPerReducer=1000000000 maxReducers=999
 totalInputFileSize=477
2020-11-25 04:42:07,624 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Setting Parallelism to 1
```

```
2020-11-25 04:42:08,616 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - creating jar file Job25487612432153817.jar
2020-11-25 04:42:22,829 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - jar file Job25487612432153817.jar created
2020-11-25 04:42:22,851 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Setting up single store job
2020-11-25 04:42:22,854 [main] INFO  org.apache.pig.data.SchemaTupleFrontend - Key [pig.schematuple] is false, will not generate code.
2020-11-25 04:42:22,854 [main] INFO  org.apache.pig.data.SchemaTupleFrontend - Starting process to move generated code to distributed cache
2020-11-25 04:42:22,854 [main] INFO  org.apache.pig.data.SchemaTupleFrontend - Setting key [pig.schematuple.classes] with classes to deserialize []
2020-11-25 04:42:22,270 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 1 map-reduce job(s) waiting for submission.
2020-11-25 04:42:22,270 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2020-11-25 04:42:22,276 [JobControl] INFO  org.apache.hadoop.yarn.client.MMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2020-11-25 04:42:22,357 [JobControl] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2020-11-25 04:42:22,720 [JobControl] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2020-11-25 04:42:22,720 [JobControl] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
2020-11-25 04:42:22,725 [JobControl] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths (combined) to process : 1
2020-11-25 04:42:22,741 [JobControl] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2020-11-25 04:42:22,741 [JobControl] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
2020-11-25 04:42:22,745 [JobControl] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths (combined) to process : 1
2020-11-25 04:42:22,829 [JobControl] INFO  org.apache.hadoop.mapreduce.JobSubmitter - number of splits:2
2020-11-25 04:42:22,998 [JobControl] INFO  org.apache.hadoop.mapreduce.JobSubmitter - Submitting tokens for job: job_1606381819383_0011
2020-11-25 04:42:23,149 [JobControl] INFO  org.apache.hadoop.yarn.client.api.impl.YarnClientImpl - Submitted application application_1606381819383_0011
2020-11-25 04:42:23,155 [JobControl] INFO  org.apache.hadoop.mapreduce.Job - The url to track the job: http://quickstart.cloudera:8888/proxy/application_1606381819383_0011/
2020-11-25 04:42:23,155 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - HadoopJobId: job_1606381819383_0011
2020-11-25 04:42:23,155 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Processing aliases dept,student,x
2020-11-25 04:42:23,156 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - detailed locations: M: dept[5,7],dept[-1,-1],x[6,4],x[6,4],student[1,10],student[-1,-1],x[6,4],x[6,4] C:  R:
2020-11-25 04:42:23,156 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - More information at: http://localhost:50030/jobdetails.jsp?jobid=job_1606381819383_0011
2020-11-25 04:42:23,272 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 0% complete
2020-11-25 04:43:43,868 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 50% complete
2020-11-25 04:44:14,829 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 100% complete
2020-11-25 04:44:14,830 [main] INFO  org.apache.pig.tools.pigstats.SimplePigStats - Script Statistics:


HadoopVersion   PigVersion     UserId    StartedAt       FinishedAt      Features
2.6.0-cdh5.12.0 0.12.0-cdh5.12.0   cloudera   2020-11-25 04:42:07   2020-11-25 04:44:14   CROSS

Success!

Job Stats (time in seconds):
JobId   Maps   Reduces MaxMapTime   MinMapTime   AvgMapTime   MedianMapTime   MaxReduceTime   MinReduceTime  AvgReduceTime   MedianReducetime        Alias  Feature Outputs
job_1606381819383_0011 2    1     43    43    43    43    25    25    25    25    dept,student,x     hdfs://quickstart.cloudera:8020/tmp/temp-1348503213/tmp1660960815,

Input(s):
Successfully read 11 records from: "/user/cloudera/trainP/dept.txt"
Successfully read 9 records from: "/user/cloudera/trainP/students.txt"

Output(s):
Successfully stored 99 records (4986 bytes) in: "hdfs://quickstart.cloudera:8020/tmp/temp-1348503213/tmp1660960815"
Counters:
Total records written : 99
Total bytes written : 4986
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_1606381819383_0011


2020-11-25 04:44:15,095 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2020-11-25 04:44:15,096 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2020-11-25 04:44:15,096 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2020-11-25 04:44:15,098 [main] INFO  org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.
2020-11-25 04:44:15,114 [main] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2020-11-25 04:44:15,114 [main] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(9,mayuresh,85.66,F,40,neha,biochemist,90000)
(9,mayuresh,85.66,F,50,rahul,chemist,80000)
(9,mayuresh,85.66,F,30,govind,scientist,60000)
(9,mayuresh,85.66,F,30,krishna,scientist,80000)
(9,mayuresh,85.66,F,30,minal,scientist,60000)
(9,mayuresh,85.66,F,20,mahesh,clerk,25000)
(9,mayuresh,85.66,F,20,manish,clerk,20000)
(9,mayuresh,85.66,F,10,siddesh,engineer,44000)
(9,mayuresh,85.66,F,10,supriya,engineer,50000)
(9,mayuresh,85.66,F,10,nikhil,engineer,45000)
(9,mayuresh,85.66,F,10,sushil,engineer,50000)
(8,neha,91.22,F,40,neha,biochemist,90000)
(8,neha,91.22,F,50,rahul,chemist,80000)
(8,neha,91.22,F,30,govind,scientist,60000)
(8,neha,91.22,F,30,krishna,scientist,80000)
(8,neha,91.22,F,30,minal,scientist,60000)
(8,neha,91.22,F,20,mahesh,clerk,25000)
(8,neha,91.22,F,20,manish,clerk,20000)

(8,neha,91.22,F,10,siddesh,engineer,44000)
(8,neha,91.22,F,10,supriya,engineer,50000)
(8,neha,91.22,F,10,nikhil,engineer,45000)
(8,neha,91.22,F,10,sushil,engineer,50000)
(7,supriya,72.21,F,40,neha,biochemist,90000)
(7,supriya,72.21,F,50,rahul,chemist,80000)
(7,supriya,72.21,F,30,govind,scientist,60000)
(7,supriya,72.21,F,30,krishna,scientist,80000)
(7,supriya,72.21,F,30,minal,scientist,60000)
(7,supriya,72.21,F,20,mahesh,clerk,25000)
(7,supriya,72.21,F,20,manish,clerk,20000)
(7,supriya,72.21,F,10,siddesh,engineer,44000)
(7,supriya,72.21,F,10,supriya,engineer,50000)
(7,supriya,72.21,F,10,nikhil,engineer,45000)
(7,supriya,72.21,F,10,sushil,engineer,50000)
(6,ankita,68.0,F,40,neha,biochemist,90000)
(6,ankita,68.0,F,50,rahul,chemist,60000)
(6,ankita,68.0,F,30,govind,scientist,60000)
(6,ankita,68.0,F,30,krishna,scientist,80000)
(6,ankita,68.0,F,30,minal,scientist,60000)
(6,ankita,68.0,F,20,mahesh,clerk,25000)
(6,ankita,68.0,F,20,manish,clerk,20000)
(6,ankita,68.0,F,10,siddesh,engineer,44000)
(6,ankita,68.0,F,10,supriya,engineer,50000)
(6,ankita,68.0,F,10,nikhil,engineer,45000)
(6,ankita,68.0,F,10,sushil,engineer,50000)
(5,sumit,62.0,F,40,neha,biochemist,90000)
(5,sumit,62.0,F,50,rahul,chemist,80000)
(5,sumit,62.0,F,30,govind,scientist,60000)
(5,sumit,62.0,F,30,krishna,scientist,80000)
(5,sumit,62.0,F,30,minal,scientist,60000)
(5,sumit,62.0,F,20,mahesh,clerk,25000)
(5,sumit,62.0,F,20,manish,clerk,20000)
(5,sumit,62.0,F,10,siddesh,engineer,44000)
(5,sumit,62.0,F,10,supriya,engineer,50000)
```

```
(5,sumit,62.0,F,10,nikhil,engineer,45000)
(5,sumit,62.0,F,10,sushil,engineer,50000)
(4,vikas,70.3,F,40,neha,biochemist,50000)
(4,vikas,70.3,F,50,rahul,chemist,80000)
(4,vikas,70.3,F,30,govind,scientist,60000)
(4,vikas,70.3,F,30,krishna,scientist,80000)
(4,vikas,70.3,F,30,minal,scientist,60000)
(4,vikas,70.3,F,20,mahesh,clerk,25000)
(4,vikas,70.3,F,20,manish,clerk,20000)
(4,vikas,70.3,F,10,siddesh,engineer,44000)
(4,vikas,70.3,F,10,supriya,engineer,50000)
(4,vikas,70.3,F,10,nikhil,engineer,45000)
(4,vikas,70.3,F,10,sushil,engineer,50000)
(3,siddhesh,80.21,F,40,neha,biochemist,50000)
(3,siddhesh,80.21,F,50,rahul,chemist,80000)
(3,siddhesh,80.21,F,30,govind,scientist,60000)
(3,siddhesh,80.21,F,30,krishna,scientist,80000)
(3,siddhesh,80.21,F,30,minal,scientist,60000)
(3,siddhesh,80.21,F,20,mahesh,clerk,25000)
(3,siddhesh,80.21,F,20,manish,clerk,20000)
(3,siddhesh,80.21,F,10,siddesh,engineer,44000)
(3,siddhesh,80.21,F,10,supriya,engineer,50000)
(3,siddhesh,80.21,F,10,nikhil,engineer,45000)
(3,siddhesh,80.21,F,10,sushil,engineer,50000)
(2,deepak,70.33,F,40,neha,biochemist,50000)
(2,deepak,70.33,F,50,rahul,chemist,80000)
(2,deepak,70.33,F,30,govind,scientist,60000)
(2,deepak,70.33,F,30,krishna,scientist,80000)
(2,deepak,70.33,F,30,minal,scientist,60000)
(2,deepak,70.33,F,20,mahesh,clerk,25000)
(2,deepak,70.33,F,20,manish,clerk,20000)
(2,deepak,70.33,F,10,siddesh,engineer,44000)
(2,deepak,70.33,F,10,supriya,engineer,50000)
(2,deepak,70.33,F,10,nikhil,engineer,45000)
(2,deepak,70.33,F,10,sushil,engineer,50000)
(1,sushil,68.2,F,40,neha,biochemist,50000)
(1,sushil,68.2,F,50,rahul,chemist,80000)
(1,sushil,68.2,F,30,govind,scientist,60000)
(1,sushil,68.2,F,30,krishna,scientist,80000)
(1,sushil,68.2,F,30,minal,scientist,60000)
(1,sushil,68.2,F,20,mahesh,clerk,25000)
(1,sushil,68.2,F,20,manish,clerk,20000)
(1,sushil,68.2,F,10,siddesh,engineer,44000)
(1,sushil,68.2,F,10,supriya,engineer,50000)
(1,sushil,68.2,F,10,nikhil,engineer,45000)
(1,sushil,68.2,F,10,sushil,engineer,50000)
grunt>
```

b. **Distinct**: This operator is used to remove the duplicate tuples in a relation. It does not preserve the original order of the contents.

```
grunt> cat /user/cloudera/trainP/dummyf.txt
1,2,2,3,4
2,4,5,65
2,4,5,65
23,44,3,2
23,44,3,2
1,2,3,4
4,3,2,1
5,6,7,8
grunt>
```

```
grunt> A = load '/user/cloudera/trainP/dummyf.txt' USING PigStorage(',') as (a1:int,a2:int,a3:int,a4:int);
2020-11-25 04:52:04,620 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2020-11-25 04:52:04,621 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
grunt> dump A;
2020-11-25 04:52:16,221 [main] INFO  org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: UNKNOWN
2020-11-25 04:52:16,222 [main] INFO  org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - {RULES ENABLED=[AddForEach, ColumnMapKeyPrune, DuplicateForEachColu
mnRewrite, GroupByConstParallelSetter, ImplicitSplitInserter, LimitOptimizer, LoadTypeCastInserter, MergeFilter, MergeForEach, NewPartitionFilterOptimizer, PushDownForE
achFlatten, PushUpFilter, SplitFilter, StreamTypeCastInserter], RULES DISABLED=[FilterLogicExpressionSimplifier, PartitionFilterOptimizer]}
2020-11-25 04:52:16,242 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File concatenation threshold: 100 optimistic? false
2020-11-25 04:52:16,245 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size before optimization: 1
2020-11-25 04:52:16,245 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size after optimization: 1
2020-11-25 04:52:16,335 [main] INFO  org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2020-11-25 04:52:16,339 [main] INFO  org.apache.pig.tools.pigstats.ScriptState - Pig script settings are added to the job
2020-11-25 04:52:16,354 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - mapred.job.reduce.markreset.buffer.percent is not
 set, set to default 0.3
2020-11-25 04:52:29,379 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - jar file Job5268544009950902D2.jar created
2020-11-25 04:52:29,408 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Setting up single store job
2020-11-25 04:52:29,401 [main] INFO  org.apache.pig.data.SchemaTupleFrontend - Key [pig.schematuple] is false, will not generate code.
2020-11-25 04:52:29,401 [main] INFO  org.apache.pig.data.SchemaTupleFrontend - Starting process to move generated code to distributed cache
2020-11-25 04:52:29,401 [main] INFO  org.apache.pig.data.SchemaTupleFrontend - Setting key [pig.schematuple.classes] with classes to deserialize []
2020-11-25 04:52:29,437 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 1 map-reduce job(s) waiting for submission.
2020-11-25 04:52:29,437 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2020-11-25 04:52:29,441 [JobControl] INFO  org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2020-11-25 04:52:29,470 [JobControl] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2020-11-25 04:52:29,824 [JobControl] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2020-11-25 04:52:29,825 [JobControl] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
2020-11-25 04:52:29,828 [JobControl] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths (combined) to process : 1
2020-11-25 04:52:29,902 [JobControl] INFO  org.apache.hadoop.mapreduce.JobSubmitter - number of splits:1
2020-11-25 04:52:30,243 [JobControl] INFO  org.apache.hadoop.mapreduce.JobSubmitter - Submitting tokens for job: job_1606301819383_0012
2020-11-25 04:52:30,379 [JobControl] INFO  org.apache.hadoop.yarn.client.api.impl.YarnClientImpl - Submitted application application_1606301819383_0012
2020-11-25 04:52:30,387 [JobControl] INFO  org.apache.hadoop.mapreduce.Job - The url to track the job: http://quickstart.cloudera:8088/proxy/application_1606301819383_0
012/
2020-11-25 04:52:30,387 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - HadoopJobId: job_1606301819383_0012
2020-11-25 04:52:30,387 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Processing aliases A
2020-11-25 04:52:30,387 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - detailed locations: M: A[7,4],A[-1,-1] C:  R:
2020-11-25 04:52:30,388 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - More information at: http://localhost:50030/jobdet
ails.jsp?jobid=job_1606301819383_0012
2020-11-25 04:52:30,541 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 0% complete
2020-11-25 04:53:23,548 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 50% complete
2020-11-25 04:53:26,214 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 100% complete
2020-11-25 04:53:26,215 [main] INFO  org.apache.pig.tools.pigstats.SimplePigStats - Script Statistics:
```

```
HadoopVersion   PigVersion      UserId  StartedAt       FinishedAt      Features
2.6.0-cdh5.12.0 0.12.0-cdh5.12.0        cloudera        2020-11-25 04:52:16     2020-11-25 04:53:26     UNKNOWN

Success!

Job Stats (time in seconds):
JobId   Maps    Reduces MaxMapTime     MinMapTime      AvgMapTime      MedianMapTime   MaxReduceTime   MinReduceTime   AvgReduceTime   MedianReducetime        Alias   F
eature  Outputs
job_1606301819383_0012  1       0       22      22      22      22      n/a     n/a     n/a     n/a     A       MAP_ONLY        hdfs://quickstart.cloudera:8020/tmp/temp
-1340583213/tmp151031464,

Input(s):
Successfully read 8 records (454 bytes) from: "/user/cloudera/trainP/dummyf.txt"

Output(s):
Successfully stored 8 records (93 bytes) in: "hdfs://quickstart.cloudera:8020/tmp/temp-1340583213/tmp151031464"

Counters:
Total records written : 8
Total bytes written : 93
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_1606301819383_0012
2020-11-25 04:53:26,464 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2020-11-25 04:53:26,465 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2020-11-25 04:53:26,465 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker.address is deprecated. Instead, use mapreduce.jobtracker.address
2020-11-25 04:53:26,466 [main] INFO  org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.
2020-11-25 04:53:26,482 [main] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2020-11-25 04:53:26,482 [main] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(1,2,2,3)
(2,4,5,65)
(2,4,5,65)
(23,44,3,2)
(23,44,3,2)
(1,2,3,4)
(4,3,2,1)
(5,6,7,8)
grunt>
```

```
grunt> z = distinct A;
grunt> dump z;
2020-11-25 04:57:03,162 [main] INFO  org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: DISTINCT
2020-11-25 04:57:03,163 [main] INFO  org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - {RULES_ENABLED=[AddForEach, ColumnMapKeyPrune, DuplicateForEachColu
mnRewrite, GroupByConstParallelSetter, ImplicitSplitInserter, LimitOptimizer, LoadTypeCastInserter, MergeFilter, MergeForEach, NewPartitionFilterOptimizer, PushDownForE
achFlatten, PushUpFilter, SplitFilter, StreamTypeCastInserter], RULES_DISABLED=[FilterLogicExpressionSimplifier, PartitionFilterOptimizer]}
2020-11-25 04:57:03,185 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File concatenation threshold: 100 optimistic? false
2020-11-25 04:57:03,214 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size before optimization: 1
2020-11-25 04:57:03,214 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size after optimization: 1
2020-11-25 04:57:03,286 [main] INFO  org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2020-11-25 04:57:03,290 [main] INFO  org.apache.pig.tools.pigstats.ScriptState - Pig script settings are added to the job
2020-11-25 04:57:03,307 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - mapred.job.reduce.markreset.buffer.percent is not
set, set to default 0.3
2020-11-25 04:57:03,308 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Reduce phase detected, estimating # of required r
educers.
2020-11-25 04:57:03,308 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Using reducer estimator: org.apache.pig.backend.h
adoop.executionengine.mapReduceLayer.InputSizeReducerEstimator
2020-11-25 04:57:03,312 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.InputSizeReducerEstimator - BytesPerReducer=1000000000 maxReducers=999
totalInputFileSize=72
2020-11-25 04:57:03,312 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Setting Parallelism to 1
2020-11-25 04:57:17,908 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Setting up single store job
2020-11-25 04:57:17,910 [main] INFO  org.apache.pig.data.SchemaTupleFrontend - Key [pig.schematuple] will not generate code.
2020-11-25 04:57:17,910 [main] INFO  org.apache.pig.data.SchemaTupleFrontend - Starting process to move generated code to distributed cache
2020-11-25 04:57:17,910 [main] INFO  org.apache.pig.data.SchemaTupleFrontend - Setting key [pig.schematuple.classes] with classes to deserialize []
2020-11-25 04:57:17,912 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Setting identity combiner class.
2020-11-25 04:57:17,953 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 1 map-reduce job(s) waiting for submission.
2020-11-25 04:57:17,954 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2020-11-25 04:57:17,969 [JobControl] INFO  org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2020-11-25 04:57:18,006 [JobControl] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2020-11-25 04:57:18,346 [JobControl] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2020-11-25 04:57:18,346 [JobControl] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
2020-11-25 04:57:18,350 [JobControl] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths (combined) to process : 1
2020-11-25 04:57:18,530 [JobControl] INFO  org.apache.hadoop.mapreduce.JobSubmitter - Submitting tokens for job: job_1606301819383_0013
2020-11-25 04:57:18,626 [JobControl] INFO  org.apache.hadoop.yarn.client.api.impl.YarnClientImpl - Submitted application application_1606301819383_0013
2020-11-25 04:57:18,632 [JobControl] INFO  org.apache.hadoop.mapreduce.Job - The url to track the job: http://quickstart.cloudera:8088/proxy/application_1606301819383_0
013/
2020-11-25 04:57:18,632 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - HadoopJobId: job_1606301819383_0013
2020-11-25 04:57:18,632 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Processing aliases A
2020-11-25 04:57:18,632 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - detailed locations: M: A[7,4],A[-1,-1] C:  R:
2020-11-25 04:57:18,632 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - More information at: http://localhost:50030/jobdet
ails.jsp?jobid=job_1606301819383_0013
2020-11-25 04:57:18,711 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 0% complete
2020-11-25 04:58:07,244 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 50% complete
2020-11-25 04:58:35,117 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 100% complete
2020-11-25 04:58:35,117 [main] INFO  org.apache.pig.tools.pigstats.SimplePigStats - Script Statistics:

HadoopVersion   PigVersion      UserId  StartedAt       FinishedAt      Features
2.6.0-cdh5.12.0 0.12.0-cdh5.12.0        cloudera        2020-11-25 04:57:03     2020-11-25 04:58:35     DISTINCT

Success!

Job Stats (time in seconds):
JobId   Maps    Reduces MaxMapTime     MinMapTime      AvgMapTime      MedianMapTime   MaxReduceTime   MinReduceTime   AvgReduceTime   MedianReducetime        Alias   F
eature  Outputs
job_1606301819383_0013  1       1       20      20      20      20      21      21      21      21      A       DISTINCT        hdfs://quickstart.cloudera:8020/tmp/temp
-1340583213/tmp-688068746,

Input(s):
Successfully read 8 records (454 bytes) from: "/user/cloudera/trainP/dummyf.txt"

Output(s):
Successfully stored 6 records (69 bytes) in: "hdfs://quickstart.cloudera:8020/tmp/temp-1340583213/tmp-688068746"

Counters:
Total records written : 6
Total bytes written : 69
Spillable Memory Manager spill count : 0
```

```
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_1606301819383_0013


2020-11-25 04:58:35,368 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2020-11-25 04:58:35,369 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2020-11-25 04:58:35,370 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2020-11-25 04:58:35,371 [main] INFO  org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.
2020-11-25 04:58:35,383 [main] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2020-11-25 04:58:35,383 [main] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(1,2,2,3)
(1,2,3,4)
(2,4,5,65)
(4,3,2,1)
(5,6,7,8)
(23,44,3,2)
grunt>
```

c. **Filter:** This operator can be used to select the required data based on some conditions

```
grunt> y = FILTER A By a2 == 2;
2020-11-25 05:02:15,830 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2020-11-25 05:02:15,831 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
grunt> dump y;
2020-11-25 05:02:20,077 [main] INFO  org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: FILTER
2020-11-25 05:02:20,078 [main] INFO  org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - {RULES ENABLED=[AddForEach, ColumnMapKeyPrune, DuplicateForEachColu
mnRewrite, GroupByConstParallelSetter, ImplicitSplitInserter, LimitOptimizer, LoadTypeCastInserter, MergeFilter, MergeForEach, NewPartitionFilterOptimizer, PushDownForE
achFlatten, PushUpFilter, SplitFilter, StreamTypeCastInserter], RULES DISABLED=[FilterLogicExpressionSimplifier, PartitionFilterOptimizer]}
2020-11-25 05:02:20,100 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File concatenation threshold: 100 optimistic? false
2020-11-25 05:02:20,103 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size before optimization: 1
2020-11-25 05:02:20,103 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size after optimization: 1
2020-11-25 05:02:20,187 [main] INFO  org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2020-11-25 05:02:20,192 [main] INFO  org.apache.pig.tools.pigstats.ScriptState - Pig script settings are added to the job
2020-11-25 05:02:20,236 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - mapred.job.reduce.markreset.buffer.percent is not
 set, set to default 0.3
2020-11-25 05:02:20,976 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - creating jar file Job4431330286582263935.jar
2020-11-25 05:02:32,784 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - jar file Job4431330286582263935.jar created
2020-11-25 05:02:32,805 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Setting up single store job
2020-11-25 05:02:32,806 [main] INFO  org.apache.pig.data.SchemaTupleFrontend - Key [pig.schematuple] is false, will not generate code.
2020-11-25 05:02:32,806 [main] INFO  org.apache.pig.data.SchemaTupleFrontend - Starting process to move generated code to distributed cache
2020-11-25 05:02:32,806 [main] INFO  org.apache.pig.data.SchemaTupleFrontend - Setting key [pig.schematuple.classes] with classes to deserialize []
2020-11-25 05:02:32,832 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 1 map-reduce job(s) waiting for submission.
2020-11-25 05:02:32,832 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2020-11-25 05:02:32,892 [JobControl] INFO  org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2020-11-25 05:02:32,929 [JobControl] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2020-11-25 05:02:33,235 [JobControl] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2020-11-25 05:02:33,235 [JobControl] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
2020-11-25 05:02:33,239 [JobControl] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths (combined) to process : 1
2020-11-25 05:02:33,453 [JobControl] INFO  org.apache.hadoop.mapreduce.JobSubmitter - Submitting tokens for job: job_1606301819383_0014
2020-11-25 05:02:33,594 [JobControl] INFO  org.apache.hadoop.yarn.client.api.impl.YarnClientImpl - Submitted application application_1606301819383_0014
2020-11-25 05:02:33,602 [JobControl] INFO  org.apache.hadoop.mapreduce.Job - The url to track the job: http://quickstart.cloudera:8088/proxy/application_1606301819383_0
014/
2020-11-25 05:02:33,603 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - HadoopJobId: job_1606301819383_0014
2020-11-25 05:02:33,603 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Processing aliases A,y
2020-11-25 05:02:33,603 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - detailed locations: M: A[7,4],A[-1,-1],y[9,4] C:
R:
2020-11-25 05:02:33,603 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - More information at: http://localhost:50030/jobdet
ails.jsp?jobid=job_1606301819383_0014
2020-11-25 05:02:33,758 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 0% complete
2020-11-25 05:03:23,470 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 50% complete
2020-11-25 05:03:29,808 [main] INFO  org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to j
ob history server
2020-11-25 05:03:31,399 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 100% complete
2020-11-25 05:03:31,400 [main] INFO  org.apache.pig.tools.pigstats.SimplePigStats - Script Statistics:

HadoopVersion  PigVersion      UserId    StartedAt        FinishedAt        Features
2.6.0-cdh5.12.0 0.12.0-cdh5.12.0  cloudera  2020-11-25 05:02:20  2020-11-25 05:03:31   FILTER

Success!

Job Stats (time in seconds):
JobId   Maps   Reduces MaxMapTime   MinMapTime   AvgMapTime   MedianMapTime  MaxReduceTime  MinReduceTime  AvgReduceTime  MedianReducetime       Alias  F
eature  Outputs
job_1606301819383_0014 1    0    20    20    20    20    n/a   n/a   n/a   n/a   A,y   MAP_ONLY    hdfs://quickstart.cloudera:8020/tmp/temp
-1348503213/tmp1348908926,

Input(s):
Successfully read 8 records (454 bytes) from: "/user/cloudera/trainP/dummyf.txt"

Output(s):
Successfully stored 2 records (22 bytes) in: "hdfs://quickstart.cloudera:8020/tmp/temp-1348503213/tmp1348908926"

Counters:
Total records written : 2
Total bytes written : 22
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_1606301819383_0014


2020-11-25 05:03:31,623 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2020-11-25 05:03:31,624 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2020-11-25 05:03:31,624 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2020-11-25 05:03:31,625 [main] INFO  org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.
2020-11-25 05:03:31,640 [main] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2020-11-25 05:03:31,648 [main] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(1,2,2,3)
(1,2,3,4)
grunt>
```

Grunt>y = FILTER dept by deptname == 'engineer';
Grunt> dump y;

```
grunt> y = FILTER dept by deptname == 'engineer')
grunt> dump y;
2020-11-25 05:06:38,207 [main] INFO  org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: FILTER
2020-11-25 05:06:38,208 [main] INFO  org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - {RULES ENABLED=[AddForEach, ColumnMapKeyPrune, DuplicateForEachColu
mnRewrite, GroupByConstParallelSetter, ImplicitSplitInserter, LimitOptimizer, LoadTypeCastInserter, MergeFilter, MergeForEach, NewPartitionFilterOptimizer, PushDownForE
achFlatten, PushUpFilter, SplitFilter, StreamTypeCastInserter], RULES DISABLED=[FilterLogicExpressionSimplifier, PartitionFilterOptimizer]}
2020-11-25 05:06:38,226 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File concatenation threshold: 100 optimistic? false
2020-11-25 05:06:38,230 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size before optimization: 1
2020-11-25 05:06:38,230 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size after optimization: 1
2020-11-25 05:06:38,551 [main] INFO  org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2020-11-25 05:06:38,555 [main] INFO  org.apache.pig.tools.pigstats.ScriptState - Pig script settings are added to the job
2020-11-25 05:06:38,572 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - mapred.job.reduce.markreset.buffer.percent is not
2020-11-25 05:06:39,836 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - creating jar file Job633882326449229631.jar
2020-11-25 05:06:51,948 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - jar file Job633882326449229631.jar created
2020-11-25 05:06:51,969 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Setting up single store job
2020-11-25 05:06:51,970 [main] INFO  org.apache.pig.data.SchemaTupleFrontend - Key [pig.schematuple] is false, will not generate code.
2020-11-25 05:06:51,970 [main] INFO  org.apache.pig.data.SchemaTupleFrontend - Starting process to move generated code to distributed cache
2020-11-25 05:06:51,970 [main] INFO  org.apache.pig.data.SchemaTupleFrontend - Setting key [pig.schematuple.classes] with classes to deserialize []
2020-11-25 05:06:52,025 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 1 map-reduce job(s) waiting for submission.
2020-11-25 05:06:52,025 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2020-11-25 05:06:52,030 [JobControl] INFO  org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2020-11-25 05:06:52,859 [JobControl] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2020-11-25 05:06:52,789 [JobControl] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2020-11-25 05:06:52,789 [JobControl] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
2020-11-25 05:06:52,792 [JobControl] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths (combined) to process : 1
2020-11-25 05:06:52,867 [JobControl] INFO  org.apache.hadoop.mapreduce.JobSubmitter - number of splits:1
2020-11-25 05:06:52,960 [JobControl] INFO  org.apache.hadoop.mapreduce.JobSubmitter - Submitting tokens for job: job_1606301819383_0015
2020-11-25 05:06:53,880 [JobControl] INFO  org.apache.hadoop.yarn.client.api.impl.YarnClientImpl - Submitted application application_1606301819383_0015
2020-11-25 05:06:53,090 [JobControl] INFO  org.apache.hadoop.mapreduce.Job - The url to track the job: http://quickstart.cloudera:8088/proxy/application_1606301819383_0
015/
2020-11-25 05:06:53,092 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - HadoopJobId: job_1606301819383_0015
2020-11-25 05:06:53,092 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Processing aliases dept,y
2020-11-25 05:06:53,092 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - detailed locations: M: dept[5,7],dept[-1,-1],y[11,
4] C:  R:
2020-11-25 05:06:53,092 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - More information at: http://localhost:50030/jobdet
ails.jsp?jobid=job_1606301819383_0015
2020-11-25 05:06:53,192 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 0% complete
2020-11-25 05:07:43,098 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 50% complete
2020-11-25 05:07:49,656 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 100% complete
2020-11-25 05:07:49,057 [main] INFO  org.apache.pig.tools.pigstats.SimplePigStats - Script Statistics:

HadoopVersion   PigVersion    UserId   StartedAt        FinishedAt       Features
2.6.0-cdh5.12.0 0.12.0-cdh5.12.0    cloudera    2020-11-25 05:06:38    2020-11-25 05:07:49    FILTER

Success!

Job Stats (time in seconds):
JobId   Maps   Reduces MaxMapTime    MinMapTIme    AvgMapTime    MedianMapTime   MaxReduceTime   MinReduceTime   AvgReduceTime   MedianReducetime        Alias   F
eature  Outputs
job_1606301819383_0015 1      0      21      21      21      21      n/a     n/a     n/a     n/a     dept,y   MAP_ONLY        hdfs://quickstart.cloudera:8020/tmp/temp
-1348503213/tmp-1562494189,

Input(s):
Successfully read 11 records (652 bytes) from: "/user/cloudera/trainP/dept.txt"

Output(s):
Successfully stored 4 records (126 bytes) in: "hdfs://quickstart.cloudera:8020/tmp/temp-1348503213/tmp-1562494189"
Counters:
Total records written : 4
Total bytes written : 126
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_1606301819383_0015


2020-11-25 05:07:49,289 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2020-11-25 05:07:49,293 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2020-11-25 05:07:49,293 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2020-11-25 05:07:49,295 [main] INFO  org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.
2020-11-25 05:07:49,315 [main] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2020-11-25 05:07:49,315 [main] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(10,sushil,engineer,50000)
(10,nikhil,engineer,45000)
(10,supriya,engineer,50000)
(10,siddesh,engineer,44000)
grunt>
```

d. **FOREACH**:- This operator is used to generate data transformation based on column data

```
grunt> X = foreach dept GENERATE id,name;
grunt> dump X;
2020-11-25 05:09:54,073 [main] INFO  org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: UNKNOWN
2020-11-25 05:09:54,875 [main] INFO  org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - {RULES ENABLED=[AddForEach, ColumnMapKeyPrune, DuplicateForEachColu
mnRewrite, GroupByConstParallelSetter, ImplicitSplitInserter, LimitOptimizer, LoadTypeCastInserter, MergeFilter, MergeForEach, NewPartitionFilterOptimizer, PushDownForE
achFlatten, PushUpFilter, SplitFilter, StreamTypeCastInserter], RULES DISABLED=[FilterLogicExpressionSimplifier, PartitionFilterOptimizer]}
2020-11-25 05:09:54,078 [main] INFO  org.apache.pig.newplan.logical.rules.ColumnPruneVisitor - Columns pruned for dept: $2, $3
2020-11-25 05:09:54,092 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File concatenation threshold: 100 optimistic? false
2020-11-25 05:09:54,095 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size before optimization: 1
2020-11-25 05:09:54,095 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size after optimization: 1
2020-11-25 05:09:54,175 [main] INFO  org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2020-11-25 05:09:54,179 [main] INFO  org.apache.pig.tools.pigstats.ScriptState - Pig script settings are added to the job
2020-11-25 05:09:54,198 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - mapred.job.reduce.markreset.buffer.percent is not
set, set to default 0.3
2020-11-25 05:09:55,726 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - creating jar file Job1724607739560467346.jar
2020-11-25 05:10:07,677 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - jar file Job1724607739560467346.jar created
2020-11-25 05:10:07,698 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Setting up single store job
2020-11-25 05:10:07,698 [main] INFO  org.apache.pig.data.SchemaTupleFrontend - Key [pig.schematuple] is false, will not generate code.
2020-11-25 05:10:07,699 [main] INFO  org.apache.pig.data.SchemaTupleFrontend - Starting process to move generated code to distributed cache
2020-11-25 05:10:07,699 [main] INFO  org.apache.pig.data.SchemaTupleFrontend - Setting key [pig.schematuple.classes] with classes to deserialize []
2020-11-25 05:10:07,733 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 1 map-reduce job(s) waiting for submission.
2020-11-25 05:10:07,733 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2020-11-25 05:10:07,736 [JobControl] INFO  org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2020-11-25 05:10:07,746 [JobControl] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2020-11-25 05:10:08,032 [JobControl] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2020-11-25 05:10:08,032 [JobControl] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
2020-11-25 05:10:08,036 [JobControl] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths (combined) to process : 1
```

```
2020-11-25 05:10:08,108 [JobControl] INFO  org.apache.hadoop.mapreduce.JobSubmitter - number of splits:1
2020-11-25 05:10:08,273 [JobControl] INFO  org.apache.hadoop.mapreduce.JobSubmitter - Submitting tokens for job: job_1606301819383_0016
2020-11-25 05:10:08,419 [JobControl] INFO  org.apache.hadoop.yarn.client.api.impl.YarnClientImpl - Submitted application application_1606301819383_0016
2020-11-25 05:10:08,427 [JobControl] INFO  org.apache.hadoop.mapreduce.Job - The url to track the job: http://quickstart.cloudera:8088/proxy/application_1606301819383_0
016/
2020-11-25 05:10:08,427 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - HadoopJobId: job_1606301819383_0016
2020-11-25 05:10:08,428 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Processing aliases X,dept
2020-11-25 05:10:08,428 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - detailed locations: M: dept[5,7],X[-1,-1] C:  R:
2020-11-25 05:10:08,428 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - More information at: http://localhost:50030/jobdet
ails.jsp?jobid=job_1606301819383_0016
2020-11-25 05:10:08,574 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 0% complete
2020-11-25 05:10:58,518 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 50% complete
2020-11-25 05:11:04,829 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 100% complete
2020-11-25 05:11:04,830 [main] INFO  org.apache.pig.tools.pigstats.SimplePigStats - Script Statistics:

HadoopVersion    PigVersion    UserId    StartedAt         FinishedAt         Features
2.6.0-cdh5.12.0 0.12.0-cdh5.12.0    cloudera    2020-11-25 05:09:54    2020-11-25 05:11:04    UNKNOWN

Success!

Job Stats (time in seconds):
JobId    Maps    Reduces MaxMapTime    MinMapTime    AvgMapTime    MedianMapTime    MaxReduceTime    MinReduceTime    AvgReduceTime    MedianReducetime        Alias  F
eature  Outputs
job_1606301819383_0016  1      0      21      21      21      21      n/a    n/a    n/a    n/a    X,dept  MAP_ONLY        hdfs://quickstart.cloudera:8020/tmp/temp
-1340583213/tmp-2007464689,

Input(s):
Successfully read 11 records (652 bytes) from: "/user/cloudera/trainP/dept.txt"

Output(s):
Successfully stored 11 records (164 bytes) in: "hdfs://quickstart.cloudera:8020/tmp/temp-1340583213/tmp-2007464689"

Counters:
Total records written : 11
Total bytes written : 164
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_1606301819383_0016


2020-11-25 05:11:04,262 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2020-11-25 05:11:04,263 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2020-11-25 05:11:04,263 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2020-11-25 05:11:04,265 [main] INFO  org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.
2020-11-25 05:11:04,279 [main] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2020-11-25 05:11:04,279 [main] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(10,sushil)
(10,nikhil)
(10,supriya)
(10,siddesh)
(20,manish)
(20,mahesh)
(30,minal)
(30,krishna)
(30,govind)
(50,rahul)
(40,neha)
grunt>
```

e. **COGROUP** operator:-

```
grunt> copyFromLocal /home/cloudera/trainingPracs/dummyfb.txt /user/cloudera/trainP
grunt> cat /user/cloudera/trainP/dummyfb.txt
1,2
2,3
3,4
5,6
65,3
3,6
4,3
23,44
grunt>
```

grunt> B = load '/user/training/dumB.txt' USING PigStorage(',') as (b1:int, b2:int);
grunt> dump B;

```
grunt> B = load '/user/cloudera/trainP/dummyfb.txt' USING PigStorage(',') as (b1:int,b2:int);
grunt> dump B;
2020-11-25 05:14:38,287 [main] INFO  org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: UNKNOWN
2020-11-25 05:14:38,289 [main] INFO  org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - {RULES_ENABLED=[AddForEach, ColumnMapKeyPrune, DuplicateForEachColu
mnRewrite, GroupByConstParallelSetter, ImplicitSplitInserter, LimitOptimizer, LoadTypeCastInserter, MergeFilter, MergeForEach, NewPartitionFilterOptimizer, PushDownForE
achFlatten, PushUpFilter, SplitFilter, StreamTypeCastInserter], RULES_DISABLED=[FilterLogicExpressionSimplifier, PartitionFilterOptimizer]}
2020-11-25 05:14:38,300 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File concatenation threshold: 100 optimistic? false
2020-11-25 05:14:38,305 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size before optimization: 1
2020-11-25 05:14:38,305 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size after optimization: 1
2020-11-25 05:14:38,378 [main] INFO  org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2020-11-25 05:14:38,381 [main] INFO  org.apache.pig.tools.pigstats.ScriptState - Pig script settings are added to the job
2020-11-25 05:14:38,380 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - mapred.job.reduce.markreset.buffer.percent is not
 set, set to default 0.3
```

```
2020-11-25 05:14:39,197 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - creating jar file Job2842863976494419055.jar
2020-11-25 05:14:52,351 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - jar file Job2842863976494419055.jar created
2020-11-25 05:14:52,377 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Setting up single store job
2020-11-25 05:14:52,381 [main] INFO  org.apache.pig.data.SchemaTupleFrontend - Key [pig.schematuple] is false, will not generate code.
2020-11-25 05:14:52,381 [main] INFO  org.apache.pig.data.SchemaTupleFrontend - Starting process to move generated code to distributed cache
2020-11-25 05:14:52,382 [main] INFO  org.apache.pig.data.SchemaTupleFrontend - Setting key [pig.schematuple.classes] with classes to deserialize []
2020-11-25 05:14:52,419 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 1 map-reduce job(s) waiting for submission.
2020-11-25 05:14:52,428 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2020-11-25 05:14:52,442 [JobControl] INFO  org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2020-11-25 05:14:52,493 [JobControl] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2020-11-25 05:14:53,287 [JobControl] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2020-11-25 05:14:53,287 [JobControl] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
2020-11-25 05:14:53,295 [JobControl] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths (combined) to process : 1
2020-11-25 05:14:53,395 [JobControl] INFO  org.apache.hadoop.mapreduce.JobSubmitter - number of splits:1
2020-11-25 05:14:53,493 [JobControl] INFO  org.apache.hadoop.mapreduce.JobSubmitter - Submitting tokens for job: job_1606301819383_0018
2020-11-25 05:14:53,603 [JobControl] INFO  org.apache.hadoop.yarn.client.api.impl.YarnClientImpl - Submitted application application_1606301819383_0018
2020-11-25 05:14:53,608 [JobControl] INFO  org.apache.hadoop.mapreduce.Job - The url to track the job: http://quickstart.cloudera:8088/proxy/application_1606301819383_0
018/
2020-11-25 05:14:53,609 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - HadoopJobId: job_1606301819383_0018
2020-11-25 05:14:53,609 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Processing aliases B
2020-11-25 05:14:53,609 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - detailed locations: M: B[14,4],B[-1,-1] C:  R:
2020-11-25 05:14:53,609 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - More information at: http://localhost:50038/jobdet
ails.jsp?jobid=job_1606301819383_0018
2020-11-25 05:14:53,718 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 0% complete
2020-11-25 05:15:43,202 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 50% complete
2020-11-25 05:15:50,077 [main] INFO  org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to j
ob history server
2020-11-25 05:15:50,927 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 100% complete
2020-11-25 05:15:50,927 [main] INFO  org.apache.pig.tools.pigstats.SimplePigStats - Script Statistics:

HadoopVersion  PigVersion  UserId  StartedAt  FinishedAt  Features
2.6.0-cdh5.12.0 0.12.0-cdh5.12.0  cloudera  2020-11-25 05:14:38  2020-11-25 05:15:50  UNKNOWN

Success!

Job Stats (time in seconds):
JobId  Maps  Reduces MaxMapTime  MinMapTime  AvgMapTime  MedianMapTime  MaxReduceTime  MinReduceTime  AvgReduceTime  MedianReducetime  Alias  F
eature  Outputs
job_1606301819383_0018  1  0  20  20  20  20  n/a  n/a  n/a  n/a  B  MAP_ONLY  hdfs://quickstart.cloudera:8020/tmp/temp
-1340503213/tmp-1062959007,

Input(s):
Successfully read 8 records (418 bytes) from: "/user/cloudera/train0/dummyfb.txt"

Output(s):
Successfully stored 8 records (63 bytes) in: "hdfs://quickstart.cloudera:8020/tmp/temp-1340503213/tmp-1062959007"

Counters:
Total records written : 8
Total bytes written : 63
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_1606301819383_0018
2020-11-25 05:15:51,515 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2020-11-25 05:15:51,516 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2020-11-25 05:15:51,516 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2020-11-25 05:15:51,542 [main] INFO  org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.
2020-11-25 05:15:51,589 [main] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2020-11-25 05:15:51,589 [main] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(1,2)
(2,3)
(3,4)
(5,6)
(65,3)
(3,6)
(4,3)
(23,44)
grunt>
```

Grunt> C = COGROUP A By a1 inner, B BY b1 Inner;
Grunt>Dump C;

```
grunt> C = COGROUP A By a1 inner, B BY b1 Inner;
grunt> dump C;
2020-11-25 05:17:40,078 [main] INFO  org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: COGROUP
2020-11-25 05:17:40,088 [main] INFO  org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - {RULES ENABLED=[AddForEach, ColumnMapKeyPrune, DuplicateForEachColu
mnRewrite, GroupByConstParallelSetter, ImplicitSplitInserter, LimitOptimizer, LoadTypeCastInserter, MergeFilter, MergeForEach, NewPartitionFilterOptimizer, PushDownForE
achFlatten, PushUpFilter, SplitFilter, StreamTypeCastInserter], RULES_DISABLED=[FilterLogicExpressionSimplifier, PartitionFilterOptimizer]}
2020-11-25 05:17:40,119 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File concatenation threshold: 100 optimistic? false
2020-11-25 05:17:40,127 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size before optimization: 1
2020-11-25 05:17:40,127 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size after optimization: 1
2020-11-25 05:17:40,211 [main] INFO  org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2020-11-25 05:17:40,215 [main] INFO  org.apache.pig.tools.pigstats.ScriptState - Pig script settings are added to the job
2020-11-25 05:17:40,271 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - mapred.job.reduce.markreset.buffer.percent is not
 set, set to default 0.3
2020-11-25 05:17:40,272 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Reduce phase detected, estimating # of required r
educers.
2020-11-25 05:17:40,275 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Using reducer estimator: org.apache.pig.backend.h
adoop.executionengine.mapReduceLayer.InputSizeReducerEstimator
2020-11-25 05:17:40,295 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.InputSizeReducerEstimator - BytesPerReducer=1000000000 maxReducers=999
 totalInputFileSize=107
2020-11-25 05:17:40,295 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Setting Parallelism to 1
```

```
2020-11-25 05:17:42.055 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - creating jar file Job3174296207502596419.jar
2020-11-25 05:17:54.300 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - jar file Job3174296207502596419.jar created
2020-11-25 05:17:54.328 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Setting up single store job
2020-11-25 05:17:54.330 [main] INFO  org.apache.pig.data.SchemaTupleFrontend - Key [pig.schematuple] is false, will not generate code.
2020-11-25 05:17:54.330 [main] INFO  org.apache.pig.data.SchemaTupleFrontend - Starting process to move generated code to distributed cache
2020-11-25 05:17:54.330 [main] INFO  org.apache.pig.data.SchemaTupleFrontend - Setting key [pig.schematuple.classes] with classes to deserialize []
2020-11-25 05:17:54.386 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 1 map-reduce job(s) waiting for submission.
2020-11-25 05:17:54.386 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2020-11-25 05:17:54.390 [JobControl] INFO  org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2020-11-25 05:17:54.422 [JobControl] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2020-11-25 05:17:55.143 [JobControl] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2020-11-25 05:17:55.144 [JobControl] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
2020-11-25 05:17:55.147 [JobControl] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths (combined) to process : 1
2020-11-25 05:17:55.179 [JobControl] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2020-11-25 05:17:55.179 [JobControl] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
2020-11-25 05:17:55.188 [JobControl] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths (combined) to process : 1
2020-11-25 05:17:55.273 [JobControl] INFO  org.apache.hadoop.mapreduce.JobSubmitter - number of splits:2
2020-11-25 05:17:55.372 [JobControl] INFO  org.apache.hadoop.mapreduce.JobSubmitter - Submitting tokens for job: job_1606301819383_0019
2020-11-25 05:17:55.517 [JobControl] INFO  org.apache.hadoop.yarn.client.api.impl.YarnClientImpl - Submitted application application_1606301819383_0019
2020-11-25 05:17:55.523 [JobControl] INFO  org.apache.hadoop.mapreduce.Job - The url to track the job: http://quickstart.cloudera:8088/proxy/application_1606301819383_0019/
2020-11-25 05:17:55.523 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - HadoopJobId: job_1606301819383_0019
2020-11-25 05:17:55.523 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Processing aliases A,B,C
2020-11-25 05:17:55.523 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - detailed locations: M: A[7,4],A[-1,-1],C[15,4],0[1,4],8[-1,-1],C[15,4] C: R:
2020-11-25 05:17:55.523 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - More information at: http://localhost:50030/jobdetails.jsp?jobid=job_1606301819383_0019
2020-11-25 05:17:55.618 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 0% complete
2020-11-25 05:19:04.210 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 50% complete
2020-11-25 05:19:31.761 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 100% complete
2020-11-25 05:19:31.762 [main] INFO  org.apache.pig.tools.pigstats.SimplePigStats - Script Statistics:
HadoopVersion   PigVersion      UserId  StartedAt       FinishedAt      Features
2.6.0-cdh5.12.0 0.12.0-cdh5.12.0        cloudera        2020-11-25 05:17:40     2020-11-25 05:19:31     COGROUP

Success!

Job Stats (time in seconds):
JobId   Maps    Reduces MaxMapTime      MinMapTime      AvgMapTime      MedianMapTime   MaxReduceTime   MinReduceTime   AvgReduceTime   MedianReducetime        Alias   Feature Outputs
job_1606301819383_0019 2        1       48      48      48      48      21      21      21      21      A,B,C   COGROUP hdfs://quickstart.cloudera:8020/tmp/temp-1340583213/tmp-117546071,

Input(s):
Successfully read 8 records from: "/user/cloudera/trainP/dummyfb.txt"
Successfully read 8 records from: "/user/cloudera/trainP/dummyf.txt"

Output(s):
Successfully stored 5 records (142 bytes) in: "hdfs://quickstart.cloudera:8020/tmp/temp-1340583213/tmp-117546071"

Counters:
Total records written : 5
Total bytes written : 142
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_1606301819383_0019
2020-11-25 05:19:32.249 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2020-11-25 05:19:32.254 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2020-11-25 05:19:32.254 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2020-11-25 05:19:32.260 [main] INFO  org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.
2020-11-25 05:19:32.290 [main] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2020-11-25 05:19:32.290 [main] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(1,{(1,2,3,4),(1,2,2,3)},{(1,2)})
(2,{(2,4,5,65),(2,4,5,65)},{(2,3)})
(4,{(4,3,2,1)},{(4,3)})
(5,{(5,6,7,8)},{(5,6)})
(23,{(23,44,3,2),(23,44,3,2)},{(23,44)})
grunt>
```

8. __Nested Projection:__

If one of the field in input relation is a [tuple,bag] then we can perform projection
operation on that field.
Grunt>Z=FOREACH C GENERATE group, B.b2;
Grunt>dump Z;

```
grunt> Z = foreach C GENERATE group, B.b2;
grunt> dump Z;
2020-11-25 05:21:28.828 [main] INFO  org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: COGROUP
2020-11-25 05:21:28.832 [main] INFO  org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - {RULES ENABLED=[AddForEach, ColumnMapKeyPrune, DuplicateForEachColumnRewrite, GroupByConstParallelSetter, ImplicitSplitInserter, LimitOptimizer, LoadTypeCastInserter, MergeFilter, MergeForEach, NewPartitionFilterOptimizer, PushDownForEachFlatten, PushUpFilter, SplitFilter, StreamTypeCastInserter], RULES DISABLED=[FilterLogicExpressionSimplifier, PartitionFilterOptimizer]}
2020-11-25 05:21:28.840 [main] INFO  org.apache.pig.newplan.logical.rules.ColumnPruneVisitor - Columns pruned for A: $1, $2, $3
2020-11-25 05:21:28.865 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File concatenation threshold: 100 optimistic? false
2020-11-25 05:21:28.870 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size before optimization: 1
2020-11-25 05:21:28.870 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size after optimization: 1
2020-11-25 05:21:28.901 [main] INFO  org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2020-11-25 05:21:28.905 [main] INFO  org.apache.pig.tools.pigstats.ScriptState - Pig script settings are added to the job
2020-11-25 05:21:29.000 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - mapred.job.reduce.markreset.buffer.percent is not set, set to default 0.3
2020-11-25 05:21:29.007 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Reduce phase detected, estimating # of required reducers.
2020-11-25 05:21:29.007 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Using reducer estimator: org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.InputSizeReducerEstimator
2020-11-25 05:21:29.015 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.InputSizeReducerEstimator - BytesPerReducer=1000000000 maxReducers=999 totalInputFileSize=107
2020-11-25 05:21:29.015 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Setting Parallelism to 1
```

```
2020-11-25 05:21:30,653 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - creating jar file Job7071870210231782646.jar
2020-11-25 05:21:42,503 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - jar file Job7071870210231782646.jar created
2020-11-25 05:21:42,523 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Setting up single store job
2020-11-25 05:21:42,525 [main] INFO  org.apache.pig.data.SchemaTupleFrontend - Key [pig.schematuple] is false, will not generate code.
2020-11-25 05:21:42,525 [main] INFO  org.apache.pig.data.SchemaTupleFrontend - Starting process to move generated code to distributed cache
2020-11-25 05:21:42,525 [main] INFO  org.apache.pig.data.SchemaTupleFrontend - Setting key [pig.schematuple.classes] with classes to deserialize []
2020-11-25 05:21:42,568 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 1 map-reduce job(s) waiting for submission.
2020-11-25 05:21:42,568 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2020-11-25 05:21:42,574 [JobControl] INFO  org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2020-11-25 05:21:42,593 [JobControl] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2020-11-25 05:21:43,258 [JobControl] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2020-11-25 05:21:43,258 [JobControl] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
2020-11-25 05:21:43,262 [JobControl] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths (combined) to process : 1
2020-11-25 05:21:43,269 [JobControl] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2020-11-25 05:21:43,270 [JobControl] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
2020-11-25 05:21:43,279 [JobControl] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths (combined) to process : 1
2020-11-25 05:21:43,351 [JobControl] INFO  org.apache.hadoop.mapreduce.JobSubmitter - number of splits:2
2020-11-25 05:21:43,543 [JobControl] INFO  org.apache.hadoop.mapreduce.JobSubmitter - Submitting tokens for job: job_1606301819383_0020
2020-11-25 05:21:43,682 [JobControl] INFO  org.apache.hadoop.yarn.client.api.impl.YarnClientImpl - Submitted application application_1606301819383_0020
2020-11-25 05:21:43,687 [JobControl] INFO  org.apache.hadoop.mapreduce.Job - The url to track the job: http://quickstart.cloudera:8088/proxy/application_1606301819383_0020/
2020-11-25 05:21:43,687 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - HadoopJobId: job_1606301819383_0020
2020-11-25 05:21:43,687 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Processing aliases A,B,C,Z
2020-11-25 05:21:43,687 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - detailed locations: M: A[7,4],A[-1,-1],C[15,4],B[14,4],B[-1,-1],C[15,4] C:  R: Z[-1,-1]
2020-11-25 05:21:43,687 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - More information at: http://localhost:50030/jobdetails.jsp?jobid=job_1606301819383_0020
2020-11-25 05:21:43,870 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 0% complete
2020-11-25 05:22:51,843 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 50% complete
2020-11-25 05:23:20,272 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 100% complete
2020-11-25 05:23:20,273 [main] INFO  org.apache.pig.tools.pigstats.SimplePigStats - Script Statistics:

HadoopVersion  PigVersion     UserId  StartedAt      FinishedAt     Features
2.6.0-cdh5.12.0 0.12.0-cdh5.12.0  cloudera  2020-11-25 05:21:28  2020-11-25 05:23:20  COGROUP

Success!

Job Stats (time in seconds):
JobId  Maps  Reduces MaxMapTime  MinMapTime  AvgMapTime  MedianMapTime  MaxReduceTime  MinReduceTime  AvgReduceTime  MedianReducetime   Alias Feature Outputs
job_1606301819383_0020 2    1     38    38    38    38    21    21    21    21    A,B,C,Z COGROUP hdfs://quickstart.cloudera:8020/tmp/temp-1340583213/tmp1784299764,

Input(s):
Successfully read 8 records from: "/user/cloudera/trainP/dummyf.txt"

Successfully read 8 records from: "/user/cloudera/trainP/dummyfb.txt"

Output(s):
Successfully stored 5 records (54 bytes) in: "hdfs://quickstart.cloudera:8020/tmp/temp-1340583213/tmp1784299764"

Counters:
Total records written : 5
Total bytes written : 54
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_1606301819383_0020


2020-11-25 05:23:20,524 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2020-11-25 05:23:20,525 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2020-11-25 05:23:20,525 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2020-11-25 05:23:20,527 [main] INFO  org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.
2020-11-25 05:23:20,541 [main] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2020-11-25 05:23:20,541 [main] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(1,{(2)})
(2,{(3)})
(4,{(3)})
(5,{(6)})
(23,{(44)})
grunt> ▮
```

Grunt>Z = FOREACH C GENERATE group,A.(a1,a2);
Grunt>dump Z;

```
grunt> Z = foreach C GENERATE group,A.(a1,a2);
grunt> dump Z;
2020-11-25 05:26:22,084 [main] INFO  org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: COGROUP
2020-11-25 05:26:22,088 [main] INFO  org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - {RULES_ENABLED=[AddForEach, ColumnMapKeyPrune, DuplicateForEachColumnRewrite, GroupByConstParallelSetter, ImplicitSplitInserter, LimitOptimizer, LoadTypeCastInserter, MergeFilter, MergeForEach, NewPartitionFilterOptimizer, PushDownForEachFlatten, PushUpFilter, SplitFilter, StreamTypeCastInserter], RULES_DISABLED=[FilterLogicExpressionSimplifier, PartitionFilterOptimizer]}
2020-11-25 05:26:22,118 [main] INFO  org.apache.pig.newplan.logical.rules.ColumnPruneVisitor - Columns pruned for B: $1
2020-11-25 05:26:22,106 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File concatenation threshold: 100 optimistic? false
2020-11-25 05:26:22,177 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size before optimization: 1
2020-11-25 05:26:22,177 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size after optimization: 1
2020-11-25 05:26:22,354 [main] INFO  org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2020-11-25 05:26:22,361 [main] INFO  org.apache.pig.tools.pigstats.ScriptState - Pig script settings are added to the job
2020-11-25 05:26:22,443 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - mapred.job.reduce.markreset.buffer.percent is not set, set to default 0.3
2020-11-25 05:26:22,445 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Reduce phase detected, estimating # of required reducers.
2020-11-25 05:26:22,445 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Using reducer estimator: org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.InputSizeReducerEstimator
2020-11-25 05:26:22,459 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.InputSizeReducerEstimator - BytesPerReducer=1000000000 maxReducers=999 totalInputFileSize=107
2020-11-25 05:26:22,459 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Setting Parallelism to 1
```

```
2020-11-25 05:26:23,513 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - creating jar file Job3400313706985982556.jar
2020-11-25 05:26:35,385 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - jar file Job3400313706985982556.jar created
2020-11-25 05:26:35,406 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Setting up single store job
2020-11-25 05:26:35,407 [main] INFO  org.apache.pig.data.SchemaTupleFrontend - Key [pig.schematuple] is false, will not generate code.
2020-11-25 05:26:35,407 [main] INFO  org.apache.pig.data.SchemaTupleFrontend - Starting process to move generated code to distributed cache
2020-11-25 05:26:35,408 [main] INFO  org.apache.pig.data.SchemaTupleFrontend - Setting key [pig.schematuple.classes] with classes to deserialize []
2020-11-25 05:26:35,449 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 1 map-reduce job(s) waiting for submission.
2020-11-25 05:26:35,450 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2020-11-25 05:26:35,454 [JobControl] INFO  org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2020-11-25 05:26:35,477 [JobControl] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2020-11-25 05:26:35,758 [JobControl] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2020-11-25 05:26:35,758 [JobControl] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
2020-11-25 05:26:35,762 [JobControl] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths (combined) to process : 1
2020-11-25 05:26:35,769 [JobControl] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2020-11-25 05:26:35,769 [JobControl] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
2020-11-25 05:26:35,775 [JobControl] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths (combined) to process : 1
2020-11-25 05:26:35,849 [JobControl] INFO  org.apache.hadoop.mapreduce.JobSubmitter - number of splits:2
2020-11-25 05:26:35,940 [JobControl] INFO  org.apache.hadoop.mapreduce.JobSubmitter - Submitting tokens for job: job_1600301819383_0021
2020-11-25 05:26:36,839 [JobControl] INFO  org.apache.hadoop.yarn.client.api.impl.YarnClientImpl - Submitted application application_1600301819383_0021
2020-11-25 05:26:36,044 [JobControl] INFO  org.apache.hadoop.mapreduce.Job - The url to track the job: http://quickstart.cloudera:8088/proxy/application_1600301819383_0021/
2020-11-25 05:26:36,044 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - HadoopJobId: job_1600301819383_0021
2020-11-25 05:26:36,044 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Processing aliases A,B,C,Z
2020-11-25 05:26:36,044 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - detailed locations: M: A[7,4],A[-1,-1],C[15,4],B[1
4,4],B[-1,-1],C[15,4] C:  R: Z[-1,-1]
2020-11-25 05:26:36,044 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - More information at: http://localhost:50030/jobdet
ails.jsp?jobid=job_1600301819383_0021
2020-11-25 05:26:36,148 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 0% complete
2020-11-25 05:27:43,299 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 25% complete
2020-11-25 05:27:43,912 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 50% complete
2020-11-25 05:28:12,106 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 100% complete
2020-11-25 05:28:12,106 [main] INFO  org.apache.pig.tools.pigstats.SimplePigStats - Script Statistics:

HadoopVersion   PigVersion       UserId   StartedAt          FinishedAt        Features
2.6.0-cdh5.12.0 0.12.0-cdh5.12.0 cloudera 2020-11-25 05:26:22 2020-11-25 05:28:12  COGROUP

Success!

Job Stats (time in seconds):
JobId   Maps    Reduces MaxMapTime      MinMapTime      AvgMapTime      MedianMapTime   MaxReduceTime   MinReduceTime   AvgReduceTime   MedianReducetime        Alias   F
eature  Outputs
job_1600301819383_0021 2     1       38      38      38      38      21      21      21      21      A,B,C,Z COGROUP hdfs://quickstart.cloudera:8020/tmp/temp-1340503
213/tmp551033165,
Input(s):
Successfully read 8 records from: "/user/cloudera/trainP/dummyf.txt"
Successfully read 8 records from: "/user/cloudera/trainP/dummyfb.txt"

Output(s):
Successfully stored 5 records (77 bytes) in: "hdfs://quickstart.cloudera:8020/tmp/temp-1340503213/tmp551033165"

Counters:
Total records written : 5
Total bytes written : 77
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_1600301819383_0021


2020-11-25 05:28:12,388 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2020-11-25 05:28:12,389 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2020-11-25 05:28:12,389 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2020-11-25 05:28:12,391 [main] INFO  org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.
2020-11-25 05:28:12,410 [main] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2020-11-25 05:28:12,411 [main] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(1,{(1,2),(1,2)})
(2,{(2,4),(2,4)})
(4,{(4,3)})
(5,{(5,6)})
(23,{(23,44),(23,44)})
grunt>
```