

# Pig Case Study

*Problem Statement:*

*What are the movie titles that the user has rated?*

*How many times a movie has been rated by the user?*

*In question 2 above, what is the average rating given for a movie?*

Loading the Data sets in to HDFS

```
[acadgild@localhost pig]$ hdfs dfs -ls /pig
18/12/23 18:43:16 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 4 items
-rw-r--r-- 1 acadgild supergroup 273 2018-12-23 15:06 /pig/employee_details.txt
-rw-r--r-- 1 acadgild supergroup 2283410 2018-12-23 18:41 /pig/movies.csv
-rw-r--r-- 1 acadgild supergroup 93 2018-12-23 14:02 /pig/pigtestfile.txt
-rw-r--r-- 1 acadgild supergroup 709550327 2018-12-23 18:42 /pig/ratings.csv
[acadgild@localhost pig]$ ll
```

Running the jar for output

Hadoop jar <Class> <file 1> <file 2> <output>

hadoop jar session7.jar com.acadgild.hadoop.session7Todo.CaseStudyIUseCasesDriver  
/pig/movies.csv /pig/ratings.csv /pig/out

```
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost pig]$ hadoop jar session7.jar com.acadgild.hadoop.session7Todo.CaseStudyIUseCasesDriver /pig/movies.csv /pig/ratings.csv /pig/out
18/12/23 18:50:27 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/12/23 18:50:29 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/12/23 18:50:31 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
18/12/23 18:50:31 INFO input.FileInputFormat: Total input paths to process : 1
18/12/23 18:50:31 INFO input.FileInputFormat: Total input paths to process : 1
18/12/23 18:50:31 INFO mapreduce.JobSubmitter: number of splits:7
18/12/23 18:50:32 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1545545220310_0017
18/12/23 18:50:32 INFO impl.YarnClientImpl: Submitted application application_1545545220310_0017
18/12/23 18:50:32 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1545545220310_0017/
18/12/23 18:50:32 INFO mapreduce.Job: Running job: job_1545545220310_0017
```

## Out Put screen prints

```
[acadgild@localhost pig]$ hadoop jar session7.jar com.acadgild.hadoop.session7Todo.CaseStudyIUseCasesDriver /pig/movies.csv /pig/ratings.csv /pig/out
19/01/08 17:51:51 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
19/01/08 17:51:52 INFO Client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
19/01/08 17:51:54 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with
ToolRunner to remedy this.
19/01/08 17:51:55 INFO input.FileInputFormat: Total input paths to process : 1
19/01/08 17:51:55 INFO input.FileInputFormat: Total input paths to process : 1
19/01/08 17:51:55 INFO mapreduce.JobSubmitter: number of splits:7
19/01/08 17:51:57 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1546949950661_0001
19/01/08 17:51:58 INFO impl.YarnClientImpl: Submitted application application_1546949950661_0001
19/01/08 17:51:58 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1546949950661_0001/
19/01/08 17:51:58 INFO mapreduce.Job: Running job: job_1546949950661_0001
19/01/08 17:52:11 INFO mapreduce.Job: Job job_1546949950661_0001 running in uber mode : false
19/01/08 17:52:11 INFO mapreduce.Job: map 0% reduce 0%
19/01/08 12:22:52 INFO mapreduce.Job: map 4% reduce 0%
19/01/08 12:22:56 INFO mapreduce.Job: map 8% reduce 0%
19/01/08 12:22:59 INFO mapreduce.Job: map 14% reduce 0%
19/01/08 12:23:02 INFO mapreduce.Job: map 19% reduce 0%
19/01/08 12:23:06 INFO mapreduce.Job: map 25% reduce 0%
19/01/08 12:23:09 INFO mapreduce.Job: map 28% reduce 0%
19/01/08 12:23:13 INFO mapreduce.Job: map 32% reduce 0%
19/01/08 12:23:16 INFO mapreduce.Job: map 33% reduce 0%
19/01/08 12:23:17 INFO mapreduce.Job: map 35% reduce 0%
19/01/08 12:23:20 INFO mapreduce.Job: map 37% reduce 0%
19/01/08 12:23:23 INFO mapreduce.Job: map 39% reduce 0%
19/01/08 12:23:26 INFO mapreduce.Job: map 40% reduce 0%
19/01/08 12:23:31 INFO mapreduce.Job: map 45% reduce 0%
19/01/08 12:23:54 INFO mapreduce.Job: map 46% reduce 0%
19/01/08 12:23:56 INFO mapreduce.Job: map 60% reduce 0%
19/01/08 12:23:57 INFO mapreduce.Job: map 63% reduce 0%
19/01/08 12:24:00 INFO mapreduce.Job: map 65% reduce 0%
19/01/08 12:24:01 INFO mapreduce.Job: map 66% reduce 0%
19/01/08 12:24:03 INFO mapreduce.Job: map 68% reduce 0%
19/01/08 12:24:04 INFO mapreduce.Job: map 70% reduce 0%
19/01/08 12:24:09 INFO mapreduce.Job: map 71% reduce 0%
19/01/08 12:24:10 INFO mapreduce.Job: map 72% reduce 0%
19/01/08 12:24:12 INFO mapreduce.Job: map 74% reduce 0%
19/01/08 12:24:13 INFO mapreduce.Job: map 76% reduce 0%
19/01/08 12:24:15 INFO mapreduce.Job: map 76% reduce 10%
19/01/08 12:24:38 INFO mapreduce.Job: map 81% reduce 10%
19/01/08 12:24:41 INFO mapreduce.Job: map 84% reduce 10%
19/01/08 12:24:44 INFO mapreduce.Job: map 87% reduce 10%
19/01/08 12:24:47 INFO mapreduce.Job: map 91% reduce 10%
```

Source: sunnort MohaXterm hu subscription to the nonfunctional edition here: <https://mohaxterm.mohatek.net>

```
19/01/08 12:25:16 INFO mapreduce.Job: map 100% reduce 29%
19/01/08 12:25:35 INFO mapreduce.Job: map 100% reduce 67%
19/01/08 12:25:44 INFO mapreduce.Job: map 100% reduce 68%
19/01/08 12:25:54 INFO mapreduce.Job: map 100% reduce 69%
19/01/08 12:26:00 INFO mapreduce.Job: map 100% reduce 70%
19/01/08 12:26:00 INFO mapreduce.Job: map 100% reduce 71%
19/01/08 12:26:18 INFO mapreduce.Job: map 100% reduce 72%
19/01/08 12:26:24 INFO mapreduce.Job: map 100% reduce 73%
19/01/08 12:26:33 INFO mapreduce.Job: map 100% reduce 74%
19/01/08 12:26:39 INFO mapreduce.Job: map 100% reduce 75%
19/01/08 12:26:52 INFO mapreduce.Job: map 100% reduce 76%
19/01/08 12:26:58 INFO mapreduce.Job: map 100% reduce 77%
19/01/08 12:27:08 INFO mapreduce.Job: map 100% reduce 78%
19/01/08 12:27:17 INFO mapreduce.Job: map 100% reduce 79%
19/01/08 12:27:27 INFO mapreduce.Job: map 100% reduce 80%
19/01/08 12:27:36 INFO mapreduce.Job: map 100% reduce 81%
19/01/08 12:27:45 INFO mapreduce.Job: map 100% reduce 82%
19/01/08 12:27:58 INFO mapreduce.Job: map 100% reduce 83%
19/01/08 12:28:07 INFO mapreduce.Job: map 100% reduce 84%
19/01/08 12:28:19 INFO mapreduce.Job: map 100% reduce 85%
19/01/08 12:28:28 INFO mapreduce.Job: map 100% reduce 86%
19/01/08 12:28:34 INFO mapreduce.Job: map 100% reduce 87%
19/01/08 12:28:43 INFO mapreduce.Job: map 100% reduce 88%
19/01/08 12:28:52 INFO mapreduce.Job: map 100% reduce 89%
19/01/08 12:29:02 INFO mapreduce.Job: map 100% reduce 90%
19/01/08 12:29:08 INFO mapreduce.Job: map 100% reduce 91%
19/01/08 12:29:18 INFO mapreduce.Job: map 100% reduce 92%
19/01/08 12:29:27 INFO mapreduce.Job: map 100% reduce 93%
19/01/08 12:29:36 INFO mapreduce.Job: map 100% reduce 94%
19/01/08 12:29:42 INFO mapreduce.Job: map 100% reduce 95%
19/01/08 12:29:51 INFO mapreduce.Job: map 100% reduce 96%
19/01/08 12:30:00 INFO mapreduce.Job: map 100% reduce 97%
19/01/08 12:30:10 INFO mapreduce.Job: map 100% reduce 98%
19/01/08 12:30:19 INFO mapreduce.Job: map 100% reduce 99%
19/01/08 12:30:25 INFO mapreduce.Job: map 100% reduce 100%
19/01/08 12:30:29 INFO mapreduce.Job: Job job_1546949950661_0001 completed successfully
19/01/08 12:30:30 INFO mapreduce.Job: Counters: 50
File System Counters
  FILE: Number of bytes read=961694264
  FILE: Number of bytes written=1457486958
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
```

```
[acadgild@localhost pig]$ hdfs dfs -cat /pig/out/part-r-00000 | head
Toy Story (1995)      66008      3.880157
GoldenEye (1995)     22534      3.431841
City Hall (1996)     4436       3.232304
Curdled (1996)       217        3.099078
"Comic 1             4.000000
Up in Smoke (1957)   3          3.666667
First Daughter (1999) 3          3.333333
"Flaw 14            3.714286
Battle of Los Angeles (2011) 44      2.522727
Jason Becker: Not Dead Yet (2012) 9      3.444444
cat: Unable to write to output stream.
[acadgild@localhost pig]$
```

## IN PIG

### CODE in PIG

```
grunt> A = load '/pig/movies.csv' USING
org.apache.pig.piggybank.storage.CSVExcelStorage(',', 'NO_MULTILINE', 'UNIX', 'SKIP_IN
PUT_HEADER');
```

```
2018-12-30 23:54:38,170 [main] INFO org.apache.hadoop.conf.Configuration.deprecation -
fs.default.name is deprecated. Instead, use fs.defaultFS
```

```
grunt> B = foreach A generate (int)$0 as movieId, (chararray)$1 as title, (chararray)$2 as
genres;
```

```
grunt> C = filter B by title is not null;
```

```
grunt> D = order C by movieId asc;
```

```
grunt> F = load '/pig/ratings.csv' USING
org.apache.pig.piggybank.storage.CSVExcelStorage(',', 'NO_MULTILINE', 'UNIX', 'SKIP_IN
PUT_HEADER');
```

```
2018-12-30 23:54:52,910 [main] INFO org.apache.hadoop.conf.Configuration.deprecation -
fs.default.name is deprecated. Instead, use fs.defaultFS
```

```
grunt> G = FOREACH F generate (int)$0 as userId, (int)$1 as movieId, (int)$2 as rating,
(int)$3 as timestamp;
```

```
grunt> H = GROUP G by movieId;
```

```
grunt> I = foreach H generate group, COUNT(G.userId);
```

```
grunt> J = filter G by (rating > 1);
```

```
grunt> K = order J by movieId asc;
```

```
grunt> L = COGROUP D by movieId, K by movieId;
```

```
grunt> M = order L by $0 asc;
```

```
grunt> N = foreach M generate group, AVG(K.rating) as cnt;
```

```
grunt> O = JOIN D by $0, N by $0;
```

```
grunt> P = foreach O generate $0, $1, $2, $3, $4;
```

```
grunt> Q = join P by $0 , I by $0;
```

```
grunt> R = foreach Q generate $1, $6, $4;
```

```
grunt> store R into /pig/out;
```

## CODE SNIPET:

```
grunt> A = load '/pig/movies.csv' USING org.apache.pig.piggybank.storage.CSVExcelStorage(',','NO_MULTILINE','UNIX','SKIP_INPUT_HEADER');
2019-01-10 00:47:08,024 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
grunt> B = foreach A generate (int)$0 as movieId, (chararray)$1 as title, (chararray)$2 as genres;
grunt> C = filter B by title is not null;
grunt> D = order C by movieId asc;
grunt> F = load '/pig/ratings.csv' USING org.apache.pig.piggybank.storage.CSVExcelStorage(',','NO_MULTILINE','UNIX','SKIP_INPUT_HEADER');
2019-01-10 00:47:09,247 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
grunt> G = FOREACH F generate (int)$0 as userId, (int)$1 as movieId, (int)$2 as rating, (int)$3 as timestamp;
grunt> H = GROUP G by movieId;
grunt> I = foreach H generate group, COUNT(G.userId);
grunt> J = filter G by (rating > 1);
grunt> K = order J by movieId asc;
grunt> L = COGROUP D by movieId, K by movieId;
grunt> M = order L by $0 asc;
grunt> N = foreach M generate group, AVG(K.rating) as cnt;
grunt> O = JOIN D by $0, N by $0;
grunt> P = foreach O generate $0, $1, $2, $3, $4;
grunt> Q = join P by $0 , I by $0;
grunt> R = foreach Q generate $1, $6, $4;
grunt> store R into '/pig/out';
2019-01-10 00:47:14,716 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2019-01-10 00:47:14,860 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.textoutputformat.separator is deprecated. Instead, use mapreduce.output
, textoutputformat.separator
2019-01-10 00:47:14,967 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: HASH_JOIN,COGROUP,GROUP_BY,ORDER_BY,FILTER
2019-01-10 00:47:15,075 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2019-01-10 00:47:15,087 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.
2019-01-10 00:47:15,195 [main] INFO org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - {RULES_ENABLED=[AddForEach, ColumnMapKeyPrune, ConstantCalculator
, GroupByConstParallelSetter, LimitOptimizer, LoadTypeCastInserter, MergeFilter, MergeForEach, PartitionFilterOptimizer, PredicatePushdownOptimizer, PushDownForEachFl
atten, PushUpFilter, SplitFilter, StreamTypeCastInserter]}
2019-01-10 00:47:15,393 [main] INFO org.apache.pig.impl.util.SpillableMemoryManager - Selected heap (Tenured Gen) of size 699872512 to monitor. collectionUsageThresh
old = 489350752, usageThreshold = 489350752
2019-01-10 00:47:15,583 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File concatenation threshold: 100 optimistic? false
2019-01-10 00:47:15,702 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.CombinerOptimizerUtil - Choosing to move algebraic foreach to combiner
2019-01-10 00:47:15,729 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.SecondaryKeyOptimizerMR - Using Secondary Key Optimization for MapRe
duce node scope-117
2019-01-10 00:47:15,730 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.SecondaryKeyOptimizerMR - Using Secondary Key Optimization for MapRe
duce node scope-175
2019-01-10 00:47:15,731 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.SecondaryKeyOptimizerMR - Using Secondary Key Optimization for MapRe
duce node scope-148
2019-01-10 00:47:15,745 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler$LastInputStreamingOptimizer - Rewrite: POPackage->POForEa
ch to POPackage(JoinPackager)
```

## Out put:

```
ch to POPackage(JoinPackager)
2019-01-10 00:47:15,745 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler$LastInputStreamingOptimizer - Rewrite: POPackage->POForEa
ch to POPackage(JoinPackager)
2019-01-10 00:47:15,772 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size before optimization: 13
2019-01-10 00:47:15,776 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - Merged 8 out of total 3 MR operators.
2019-01-10 00:47:15,777 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - Merged 1 map-only splittes.
2019-01-10 00:47:15,777 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - Merged 1 map-reduce splittes.
2019-01-10 00:47:15,777 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - Merged 2 out of total 3 MR operators.
2019-01-10 00:47:15,777 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size after optimization: 11
2019-01-10 00:47:15,924 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2019-01-10 00:47:16,021 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2019-01-10 00:47:16,055 [main] INFO org.apache.pig.tools.pigstats.mapreduce.MRScriptState - Pig script settings are added to the job
2019-01-10 00:47:16,096 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.reduce.markreset.buffer.percent is deprecated. Instead, use mapred
uce.reduce.markreset.buffer.percent
2019-01-10 00:47:16,896 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - mapred.job.reduce.markreset.buffer.percent is n
ot set, set to default 0.3
2019-01-10 00:47:16,896 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.output.compress is deprecated. Instead, use mapreduce.output.fileoutput
format.compress
2019-01-10 00:47:16,902 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - This job cannot be converted run in-process
2019-01-10 00:47:16,943 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.submit.replication is deprecated. Instead, use mapreduce.client.submit.
file.replication
2019-01-10 00:47:17,728 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar file:/home/acadgild/install/pig/pig-0
.16.0/lib/piggybank.jar to DistributedCache through /tmp/tmp926958319/tmp847959287/piggybank.jar
2019-01-10 00:47:17,919 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar file:/home/acadgild/install/pig/pig-0
.16.0/pig-0.16.0-core-h2.jar to DistributedCache through /tmp/tmp926958319/tmp1858109128/pig-0.16.0-core-h2.jar
2019-01-10 00:47:17,980 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar file:/home/acadgild/install/pig/pig-0
.16.0/lib/automaton-1.11-8.jar to DistributedCache through /tmp/tmp926958319/tmp-1904668275/automaton-1.11-8.jar
2019-01-10 00:47:18,042 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar file:/home/acadgild/install/pig/pig-0
.16.0/lib/antlr-runtime-3.4.jar to DistributedCache through /tmp/tmp926958319/tmp-1590405950/antlr-runtime-3.4.jar
2019-01-10 00:47:18,099 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar file:/home/acadgild/install/pig/pig-0
.16.0/lib/joda-time-2.9.3.jar to DistributedCache through /tmp/tmp926958319/tmp085546027/joda-time-2.9.3.jar
2019-01-10 00:47:18,226 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Setting up single store job
2019-01-10 00:47:18,262 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Key [pig.schematuple] is false, will not generate code.
2019-01-10 00:47:18,262 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Starting process to move generated code to distributed cacche
2019-01-10 00:47:18,262 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Setting key [pig.schematuple.classes] with classes to deserialize []
2019-01-10 00:47:18,643 [main] INFO org.apache.pig.tools.pigstats.mapreduce.MRScriptState - Pig script settings are added to the job
2019-01-10 00:47:18,650 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - mapred.job.reduce.markreset.buffer.percent is n
ot set, set to default 0.3
2019-01-10 00:47:18,653 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Reduce phase detected, estimating # of required
reducers
2019-01-10 00:47:18,659 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Using reducer estimator: org.apache.pig.backend
.hadoop.executionengine.mapReduceLayer.InputSizeReducerEstimator
2019-01-10 00:47:18,670 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.InputSizeReducerEstimator - BytesPerReducer=1000000000 maxReducers=9
```

```

r(String).
2019-01-10 00:04:34,045 [JobControl] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2019-01-10 00:04:34,045 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
2019-01-10 00:04:34,207 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths (combined) to process : 1
2019-01-10 00:04:35,164 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter - number of splits:1
2019-01-10 00:04:35,945 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter - Submitting tokens for job: job_1547058466818_0001
2019-01-10 00:04:36,983 [JobControl] INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl - Submitted application application_1547058466818_0001
2019-01-10 00:04:37,626 [JobControl] INFO org.apache.hadoop.mapreduce.Job - The url to track the job: http://localhost:8088/proxy/application_1547058466818_0001/
2019-01-10 00:04:37,631 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - HadoopJobId: job_1547058466818_0001
2019-01-10 00:04:37,632 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Processing aliases A,B,C
2019-01-10 00:04:37,634 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - detailed locations: M: A[1,4],B[2,4],C[3,4] C:
R:
2019-01-10 00:04:37,640 [JobControl] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2019-01-10 00:04:37,763 [JobControl] WARN org.apache.hadoop.mapreduce.JobResourceUploader - No job jar file set. User classes may not be found. See Job or Job#setJ
r(String).
2019-01-10 00:04:37,867 [JobControl] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2019-01-10 00:04:37,867 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
2019-01-10 00:04:37,873 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths (combined) to process : 6
2019-01-10 00:04:38,037 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter - number of splits:6
2019-01-10 00:04:38,285 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter - Submitting tokens for job: job_1547058466818_0002
2019-01-10 00:04:38,315 [JobControl] INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl - Submitted application application_1547058466818_0002
2019-01-10 00:04:38,496 [JobControl] INFO org.apache.hadoop.mapred.YARNRunner - Job jar is not present. Not adding any jar to the list of resources.
2019-01-10 00:04:38,514 [JobControl] INFO org.apache.hadoop.mapreduce.Job - The url to track the job: http://localhost:8088/proxy/application_1547058466818_0002/
2019-01-10 00:04:38,514 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - HadoopJobId: job_1547058466818_0002
2019-01-10 00:04:38,514 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Processing aliases F,G,H,I,J
2019-01-10 00:04:38,514 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - detailed locations: M: F[5,4],G[6,4],J[9,4],I[8
4],H[7,4] C: I[8,4],H[7,4] R: I[8,4]
2019-01-10 00:04:38,625 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 0% complete
2019-01-10 00:04:38,625 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Running jobs are [job_1547058466818_0001,job_15
7058466818_0002]
2019-01-10 00:05:28,593 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 4% complete
2019-01-10 00:05:28,593 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Running jobs are [job_1547058466818_0001,job_15
7058466818_0002]
2019-01-10 00:05:34,630 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 9% complete
2019-01-10 00:05:34,630 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Running jobs are [job_1547058466818_0002]
2019-01-10 00:05:34,660 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2019-01-10 00:05:34,677 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to
job history server
2019-01-10 00:05:41,098 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2019-01-10 00:05:41,159 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to
job history server
2019-01-10 00:05:41,526 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032

```

<https://mohsatharm.mohsatharm.net>

## Result of the output

```

[acadvil@localhost pig]$ hdfs dfs -cat /pig/out/part-r-00000 | head
19/01/08 12:37:22 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Toy Story (1995)      66008      3.888157
GoldenEye (1995)     32534      3.431841
City Hall (1996)     4436       3.232384
Curdled (1996)       217        3.099078
"Comic 1             4.000000
Up in Smoke (1957)   3           3.666667
First Daughter (1999) 3           3.333333
"Flaw 14             3.714286
Battle of Los Angeles (2011) 44          2.522727
Jason Becker: Not Dead Yet (2012) 9           3.444444
cat: Unable to write to output stream.
[acadvil@localhost pig]$

```