

To-do

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?

Assignment as described above was to work with movie dataset and generate pig script and run it to generate required output.

PIG Script

```
a = load '/home/acadgild/input/movies.csv' using PigStorage(',') as (movieid:int, title:chararray, genres:chararray);
```

```
a1 = filter a by not title matches 'title*';
```

```
b = load '/home/acadgild/input/ratings.csv' using PigStorage(',') as (userId:int,movieId:int,rating:float,timestamp:bytearray);
```

```
c = group b by movieId;
```

```
d = foreach c generate group as movieId, COUNT(b.userId) as RatingUserPerMovieCount, SUM(b.rating)/COUNT(b.userId) as AverageRatingPerMovie;
```

```
store d into '/home/acadgild/output/MovieRatingAssignment.txt' using PigStorage(',');
```

```
e = load '/home/acadgild/output/MovieRatingAssignment.txt/part-r-00000' using PigStorage(',') as (MovieID:int, RatingUserPerMovieCount:int, AverageRatingPerMovie:float);
```

```
e1 = filter e by not (MovieID == 0);
```

```
f = join a1 by $0, e1 by $0;
```

```
g = foreach f generate $1, $4, $5;
```

```
store g into '/home/acadgild/output/FinalOutputAssignment.txt' using PigStorage(',');
```

After running pig script successfully

```
acadgild@localhost:~  
File Edit View Search Terminal Help  
2.6.5 0.16.0 acadgild 2018-10-18 00:33:49 2018-10-18 00:33:55 HASH_JOIN, FILTER  
  
Success!  
  
Job Stats (time in seconds):  
JobId Maps Reduces MaxMapTime MinMapTime AvgMapTime MedianMapTime MaxReduceTime MinReduceTime AvgRe  
duceTime MedianReductime Alias Feature Outputs  
job_local1042013575_0009 2 1 n/a n/a n/a n/a n/a n/a a,a1,e,e1,f,g  
HASH_JOIN /home/acadgild/output/FinalOutputAssignment.txt,  
  
Input(s):  
Successfully read 45116 records from: "/home/acadgild/output/MovieRatingAssignment.txt/part-r-00000"  
Successfully read 45844 records from: "/home/acadgild/input/movies.csv"  
  
Output(s):  
Successfully stored 45115 records in: "/home/acadgild/output/FinalOutputAssignment.txt"  
  
Counters:  
Total records written : 45115  
Total bytes written : 0  
Spillable Memory Manager spill count : 0  
Total bags proactively spilled: 0  
Total records proactively spilled: 0  
  
Job DAG:  
job_local1042013575_0009  
  
2018-10-18 00:33:55,769 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processNam  
e=JobTracker, sessionId= - already initialized  
2018-10-18 00:33:55,770 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processNam  
e=JobTracker, sessionId= - already initialized  
2018-10-18 00:33:55,771 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processNam  
e=JobTracker, sessionId= - already initialized  
2018-10-18 00:33:55,791 [main] WARN org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Encount  
ered Warning FIELD DISCARDED_TYPE CONVERSION FAILED 1 time(s).  
2018-10-18 00:33:55,791 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success  
!  
grunt> █  
acacgild@localhost:~ PigAssignmentMovieR... acadgild@localhost:~/... FinalOutputAssignmen...
```

Post running script final output created displaying head of the file

```
acadgild@localhost: ~  
File Edit View Search Terminal Help  
[acadgild@localhost ~]$ cat /home/acadgild/output/FinalOutputAssignment.txt/part-r-00000 | head  
Toy Story (1995),66008,3.8881576  
Jumanji (1995),26060,3.2369533  
Grumpier Old Men (1995),15497,3.1755502  
Waiting to Exhale (1995),2981,2.8757129  
Father of the Bride Part II (1995),15258,3.0795648  
Heat (1995),27895,3.8417637  
Sabrina (1995),15157,3.3721054  
Tom and Huck (1995),1521,3.1291914  
Sudden Death (1995),4423,3.0083654  
GoldenEye (1995),32534,3.4318404  
[acadgild@localhost ~]$
```

I

acacgild@localhost: ~ PigAssignmentMovieR... acadgild@localhost: ~/... FinalOutputAssignmen...