

## Setup Using Pig

```
grunt>
grunt> describe movies;
movies: {movieid: int,title: chararray,genres: chararray}
grunt>
grunt>
grunt>
grunt>
grunt> describe ratings;
ratings: {userid: int,movieid: int,rating: float,timestamp: int}
grunt>
```

```
grunt>
grunt> Ratings2Movies = JOIN ratings by movieid, movies by movieid;
grunt>
grunt>
grunt> describe Ratings2Movies;
Ratings2Movies: {ratings::userid: int,ratings::movieid: int,ratings::rating: flo
at,ratings::timestamp: int,movies::movieid: int,movies::title: chararray,movies:
:genres: chararray}
grunt>
```

## List of Movie Titles User 1 has Rated

```
grunt>
grunt> User1Filter = FILTER Ratings2Movies by userid==1;
```

## Results

```

2018-09-29 00:21:41,067 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2018-09-29 00:21:41,067 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(1,110,1.0,1425941529,110,Braveheart (1995),Action|Drama|War)
(1,147,4.5,1425942435,147,"Basketball Diaries, The (1995)")
(1,858,5.0,1425941523,858,"Godfather, The (1972)")
(1,1221,5.0,1425941546,1221,"Godfather: Part II, The (1974)")
(1,1246,5.0,1425941556,1246,Dead Poets Society (1989),Drama)
(1,1968,4.0,1425942148,1968,"Breakfast Club, The (1985)")
(1,2762,4.5,1425941300,2762,"Sixth Sense, The (1999)")
(1,2918,5.0,1425941593,2918,Ferris Bueller's Day Off (1986),Comedy)
(1,2959,4.0,1425941601,2959,Fight Club (1999),Action|Crime|Drama|Thriller)
(1,4226,4.0,1425942228,4226,Memento (2000),Mystery|Thriller)
(1,4878,5.0,1425941434,4878,Donnie Darko (2001),Drama|Mystery|Sci-Fi|Thriller)
(1,5577,5.0,1425941397,5577,Igby Goes Down (2002),Comedy|Drama)
(1,33794,4.0,1425942005,33794,Batman Begins (2005),Action|Crime|IMAX)
(1,54503,3.5,1425941313,54503,Superbad (2007),Comedy)
(1,58559,4.0,1425942007,58559,"Dark Knight, The (2008)")
(1,59315,5.0,1425941502,59315,Iron Man (2008),Action|Adventure|Sci-Fi)
(1,68358,5.0,1425941464,68358,Star Trek (2009),Action|Adventure|Sci-Fi|IMAX)
(1,69844,5.0,1425942139,69844,Harry Potter and the Half-Blood Prince (2009),Adventure|Fantasy|Mystery|Romance|IMAX)
(1,73017,5.0,1425942699,73017,Sherlock Holmes (2009),Action|Crime|Mystery|Thriller)
(1,81834,5.0,1425942133,81834,Harry Potter and the Deathly Hallows: Part 1 (2010),Action|Adventure|Fantasy|IMAX)
(1,91500,2.5,1425942647,91500,The Hunger Games (2012),Action|Adventure|Drama|Sci-Fi|Thriller)
(1,91542,5.0,1425942618,91542,Sherlock Holmes: A Game of Shadows (2011),Action|Adventure|Comedy|Crime|Mystery|Thriller)
(1,92439,5.0,1425941424,92439,"Art of Getting By, The (2011)")
(1,96821,5.0,1425941382,96821,"Perks of Being a Wallflower, The (2012)")
(1,98809,0.5,1425942640,98809,"Hobbit: An Unexpected Journey, The (2012)")
(1,99114,4.0,1425941667,99114,Django Unchained (2012),Action|Drama|Western)
(1,112552,5.0,1425941336,112552,Whiplash (2014),Drama)
grunt>

```

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

Created filter for Toy Story

```

grunt> filterToy = filter jRatings2Movies by rmovieid==1;

```

```

grunt> cToy = foreach filterToy Generate COUNT(ratings::userid);

```

```

grunt> gToy = group cToy all;

```

```
grunt> cToy = foreach gToy GENERATE COUNT(filterToy);
```

```
grunt> dump cToy;
```

Toy Story = 66008 times

```
Job DAG:
job_local1579593329_0005      ->      job_local769950553_0006,
job_local769950553_0006

2018-09-29 01:32:55,572 [main] INFO  org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot in
italize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-09-29 01:32:55,573 [main] INFO  org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot in
italize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-09-29 01:32:55,573 [main] INFO  org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot in
italize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-09-29 01:32:55,582 [main] INFO  org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot in
italize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-09-29 01:32:55,582 [main] INFO  org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot in
italize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-09-29 01:32:55,582 [main] INFO  org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot in
italize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-09-29 01:32:55,583 [main] WARN  org.apache.pig.backend.hadoop.executionengine.mapRed
uceLayer.MapReduceLauncher - Encountered Warning FIELD_DISCARDED_TYPE_CONVERSION_FAILED 5
time(s).
2018-09-29 01:32:55,583 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapRed
uceLayer.MapReduceLauncher - Success!
2018-09-29 01:32:55,583 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - i
o.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2018-09-29 01:32:55,586 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - f
s.default.name is deprecated. Instead, use fs.defaultFS
2018-09-29 01:32:55,586 [main] WARN  org.apache.pig.data.SchemaTupleBackend - SchemaTuple
Backend has already been initialized
2018-09-29 01:32:55,594 [main] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputForma
t - Total input paths to process : 1
2018-09-29 01:32:55,594 [main] INFO  org.apache.pig.backend.hadoop.executionengine.util.M
apRedUtil - Total input paths to process : 1
(66008)
grunt>
```

What is the average rating giving for a movie?

```

(65317,1,3.0,1068019081)
(65326,1,3.0,1167689659)
(65337,1,5.0,829413881)
(65341,1,2.0,1180813927)
(65352,1,4.5,1444747757)
grunt> gToy = Group fRatings all;
2018-09-29 02:08:15,944 [main] ERROR org.apache.pig.tools.grunt.Grunt - ERROR 1200: Pig script failed to parse:
<line 13, column 13> Undefined alias: fRatings
Details at logfile: /home/acadgild/tony/Session7/Assignment2/pig_1538162855515.log
grunt> gToy = Group fRatingsToy all;
grunt>
grunt>
grunt> describe ratings;
ratings: {userid: int,rmovieid: int,rating: float,timestamp: int}
grunt>
grunt>
grunt>
grunt> AvgToy = foreach gToy Generate (ratings.rmovieid, ratings.rating), AVG(ratings.rating);
2018-09-29 02:10:35,676 [main] ERROR org.apache.pig.tools.grunt.Grunt - ERROR 1045:
<line 14, column 67> Could not infer the matching function for org.apache.pig.builtin.AVG as multiple or none of them fit.
Please use an explicit cast.
Details at logfile: /home/acadgild/tony/Session7/Assignment2/pig_1538162855515.log
grunt>
grunt>
grunt>
grunt> AvgToy = foreach gToy Generate AVG(ratings.rating);
2018-09-29 02:12:47,621 [main] ERROR org.apache.pig.tools.grunt.Grunt - ERROR 1045:
<line 14, column 31> Could not infer the matching function for org.apache.pig.builtin.AVG as multiple or none of them fit.
Please use an explicit cast.
Details at logfile: /home/acadgild/tony/Session7/Assignment2/pig_1538162855515.log
grunt> AvgToy = foreach gToy Generate AVG(fRatingsToy.rating);
grunt>
grunt>

```

Averaging rating for Toy Story: 3.88

```

2018-09-29 02:15:25,546 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process
1
2018-09-29 02:15:25,546 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to
rocess : 1
(3.8881574960610834)
grunt>

```