

# Hive Case Study

---

Student Name: Subbham Vishal

Course: Big Data Hadoop & Spark Training

Assignment – Problem Statement:

What are the movie titles that the user has rated?

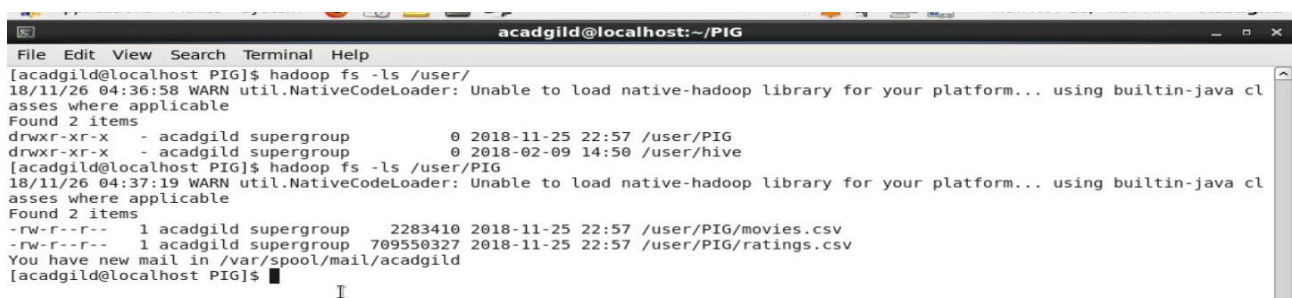
How many times a movie has been rated by the user?

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

## Case study 1

All the screenshots of the task are as follows

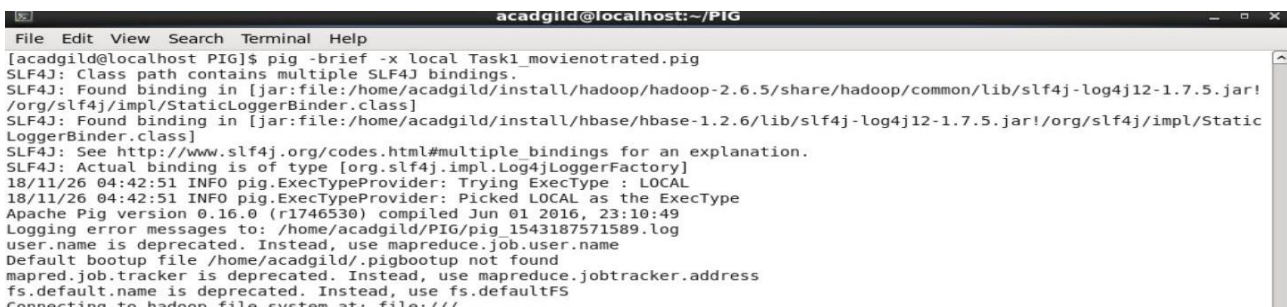
1. Place the file in Hdfs as well as the local file system.



```
acadgild@localhost: ~/PIG
File Edit View Search Terminal Help
[acadgild@localhost PIG]$ hadoop fs -ls /user/
18/11/26 04:36:58 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
Found 2 items
drwxr-xr-x - acadgild supergroup          0 2018-11-25 22:57 /user/PIG
drwxr-xr-x - acadgild supergroup          0 2018-02-09 14:50 /user/hive
[acadgild@localhost PIG]$ hadoop fs -ls /user/PIG
18/11/26 04:37:19 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
Found 2 items
-rw-r--r-- 1 acadgild supergroup 2283410 2018-11-25 22:57 /user/PIG/movies.csv
-rw-r--r-- 1 acadgild supergroup 709550327 2018-11-25 22:57 /user/PIG/ratings.csv
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost PIG]$
```

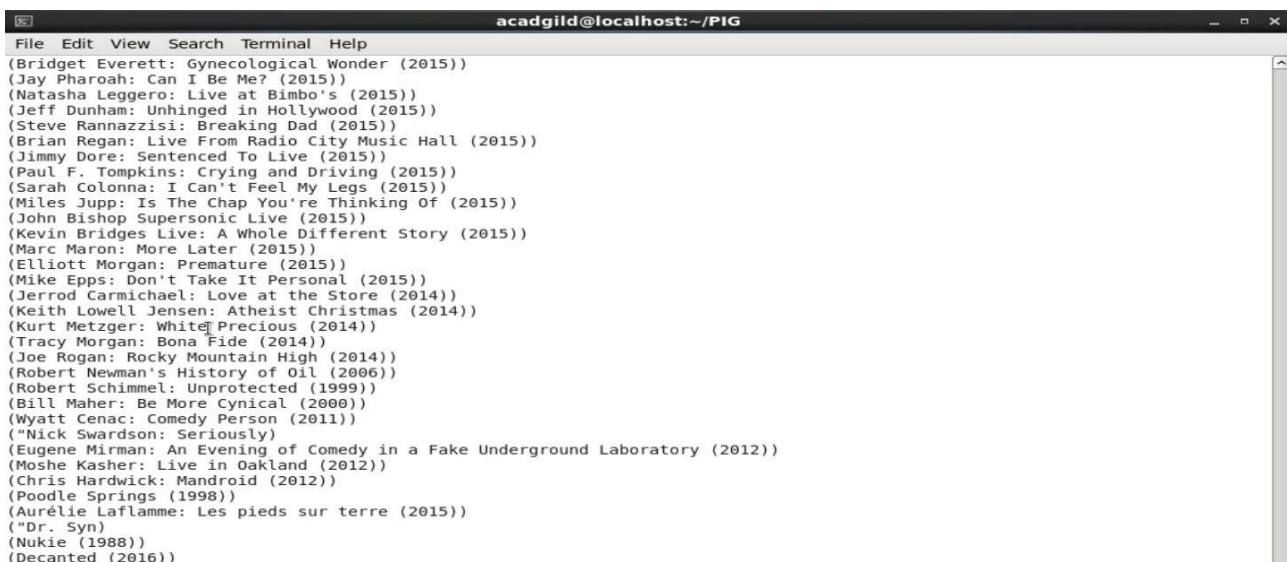
## 2. Task1

```
movie = load '/home/acadgild/PIG/movies.csv' using PigStorage(',') as (movieid:int,
title:chararray, genres:chararray);
rating = load '/home/acadgild/PIG/ratings.csv' using PigStorage(',') as (userid:int, movieid:int,
rating:int, timestamp:int);
movierating = COGROUP movie by movieid, rating by movieid;
movienotrated = filter movierating by IsEmpty(rating);
movienotratedtitle = foreach movienotrated generate flatten(movie.title);
dump movienotratedtitle;
```

A terminal window titled 'acadgild@localhost:~/PIG' showing the execution of a Pig script. The command is 'pig -brief -x local Task1\_movienotrated.pig'. The output includes SLF4J warnings about multiple bindings, version information for Pig (0.16.0), and deprecation warnings for 'user.name', 'mapred.job.tracker', and 'fs.default.name'. The script is connecting to the Hadoop file system.

```
acadgild@localhost:~/PIG
File Edit View Search Terminal Help
[acadgild@localhost PIG]$ pig -brief -x local Task1_movienotrated.pig
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/common/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/home/acadgild/install/hbase/hbase-1.2.6/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
18/11/26 04:42:51 INFO pig.ExecTypeProvider: Trying ExecType : LOCAL
18/11/26 04:42:51 INFO pig.ExecTypeProvider: Picked LOCAL as the ExecType
Apache Pig version 0.16.0 (r1746530) compiled Jun 01 2016, 23:10:49
Logging error messages to: /home/acadgild/PIG/pig_1543187571589.log
user.name is deprecated. Instead, use mapreduce.job.user.name
Default bootup file /home/acadgild/.pigbootup not found
mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
fs.default.name is deprecated. Instead, use fs.defaultFS
Connecting to hadoop file system at: file:///
```

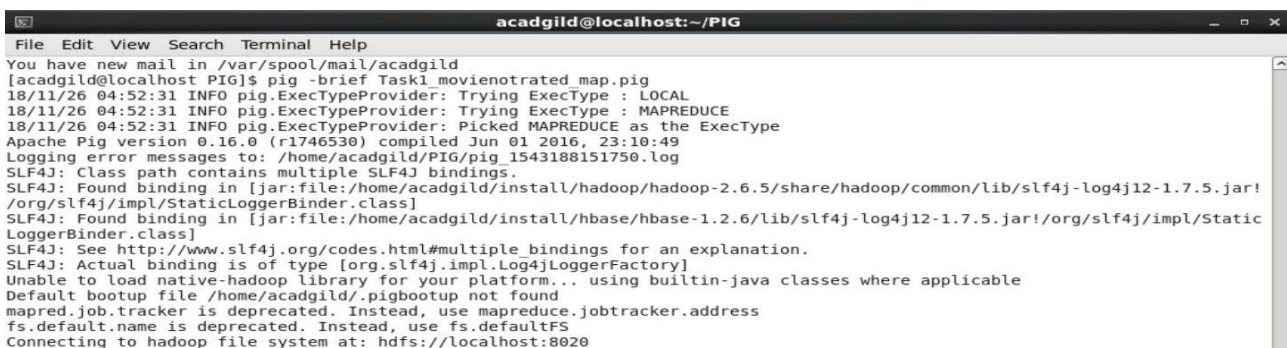
## Output

A terminal window titled 'acadgild@localhost:~/PIG' showing the output of the Pig script. The output is a list of movie titles, each followed by its year in parentheses, representing the titles of movies that were not rated in the dataset.

```
acadgild@localhost:~/PIG
File Edit View Search Terminal Help
(Bridget Everett: Gynecological Wonder (2015))
(Jay Pharoah: Can I Be Me? (2015))
(Natasha Leggero: Live at Bimbo's (2015))
(Jeff Dunham: Unhinged in Hollywood (2015))
(Steve Rannazzisi: Breaking Dad (2015))
(Brian Regan: Live From Radio City Music Hall (2015))
(Jimmy Dore: Sentenced To Live (2015))
(Paul F. Tompkins: Crying and Driving (2015))
(Sarah Colonna: I Can't Feel My Legs (2015))
(Miles Jupp: Is The Chap You're Thinking Of (2015))
(John Bishop Supersonic Live (2015))
(Kevin Bridges Live: A Whole Different Story (2015))
(Marc Maron: More Later (2015))
(Elliott Morgan: Premature (2015))
(Mike Epps: Don't Take It Personal (2015))
(Jerrold Carmichael: Love at the Store (2014))
(Keith Lowell Jensen: Atheist Christmas (2014))
(Kurt Metzger: White Precious (2014))
(Tracy Morgan: Bona Fide (2014))
(Joe Rogan: Rocky Mountain High (2014))
(Robert Newman's History of Oil (2006))
(Robert Schimmel: Unprotected (1999))
(Bill Maher: Be More Cynical (2000))
(Wyatt Cenac: Comedy Person (2011))
("Nick Swardson: Seriously)
(Eugene Mirman: An Evening of Comedy in a Fake Underground Laboratory (2012))
(Moshe Kasher: Live in Oakland (2012))
(Chris Hardwick: Mandroid (2012))
(Poodle Springs (1998))
(Aur lie Laflamme: Les pieds sur terre (2015))
("Dr. Syn)
(Nukie (1988))
(Decanted (2016))
```

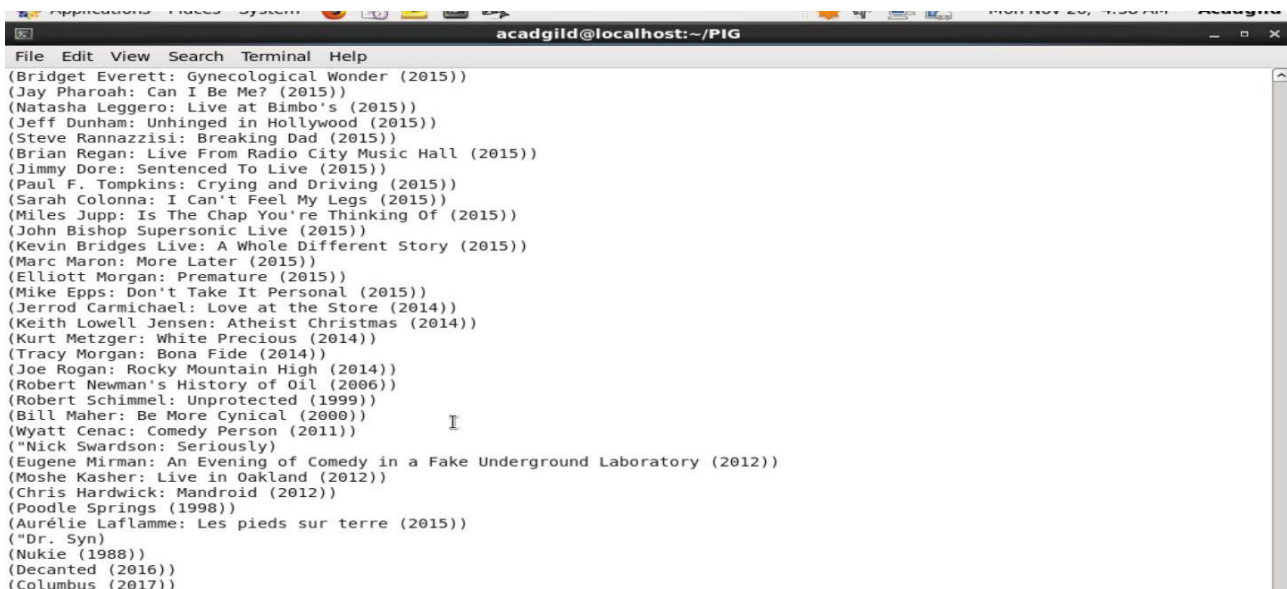
## Task2

```
movie = load 'hdfs://localhost:8020/user/PIG/movies.csv' using PigStorage(',') as
(movieid:int, title:chararray, genres:chararray);
rating = load 'hdfs://localhost:8020/user/PIG/ratings.csv' using PigStorage(',') as (userid:int,
movieid:int, rating:int, timestamp:int);
movierating = COGROUP movie by movieid, rating by movieid;
movienotrated = filter movierating by IsEmpty(rating);
movienotratedtitle = foreach movienotrated generate flatten(movie.title);
dump movienotratedtitle;
```

A terminal window titled 'acadgild@localhost:~/PIG' showing the output of a Pig script. The output includes version information, logging messages, and warnings about deprecated features. The script is 'Task1 movienotrated map.pig'.

```
acadgild@localhost:~/PIG
File Edit View Search Terminal Help
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost PIG]$ pig -brief Task1 movienotrated map.pig
18/11/26 04:52:31 INFO pig.ExecTypeProvider: Trying ExecType : LOCAL
18/11/26 04:52:31 INFO pig.ExecTypeProvider: Trying ExecType : MAPREDUCE
18/11/26 04:52:31 INFO pig.ExecTypeProvider: Picked MAPREDUCE as the ExecType
Apache Pig version 0.16.0 (r1746530) compiled Jun 01 2016, 23:10:49
Logging error messages to: /home/acadgild/PIG/pig_1543188151750.log
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/common/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/home/acadgild/install/hbase/hbase-1.2.6/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Default bootup file /home/acadgild/.pigbootup not found
mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
fs.default.name is deprecated. Instead, use fs.defaultFS
Connecting to hadoop file system at: hdfs://localhost:8020
```

## Output

A terminal window titled 'acadgild@localhost:~/PIG' showing the output of the Pig script. The output is a list of movie titles and years, sorted by rating.

```
acadgild@localhost:~/PIG
File Edit View Search Terminal Help
(Bridget Everett: Gynecological Wonder (2015))
(Jay Pharoah: Can I Be Me? (2015))
(Natasha Leggero: Live at Bimbo's (2015))
(Jeff Dunham: Unhinged in Hollywood (2015))
(Steve Rannazzisi: Breaking Dad (2015))
(Brian Regan: Live From Radio City Music Hall (2015))
(Jimmy Dore: Sentenced To Live (2015))
(Paul F. Tompkins: Crying and Driving (2015))
(Sarah Colonna: I Can't Feel My Legs (2015))
(Miles Jupp: Is The Chap You're Thinking Of (2015))
(John Bishop Supersonic Live (2015))
(Kevin Bridges Live: A Whole Different Story (2015))
(Marc Maron: More Later (2015))
(Elliott Morgan: Premature (2015))
(Mike Epps: Don't Take It Personal (2015))
(Jerrold Carmichael: Love at the Store (2014))
(Keith Lowell Jensen: Atheist Christmas (2014))
(Kurt Metzger: White Precious (2014))
(Tracy Morgan: Bona Fide (2014))
(Joe Rogan: Rocky Mountain High (2014))
(Robert Newman's History of Oil (2006))
(Robert Schimmel: Unprotected (1999))
(Bill Maher: Be More Cynical (2000))
(Wyatt Cenac: Comedy Person (2011))
("Nick Swardson: Seriously)
(Eugene Mirman: An Evening of Comedy in a Fake Underground Laboratory (2012))
(Moshe Kasher: Live in Oakland (2012))
(Chris Hardwick: Mandroid (2012))
(Poodle Springs (1998))
(Aur lie Laflamme: Les pieds sur terre (2015))
("Dr. Syn)
(Nukie (1988))
(Decanted (2016))
(Columbus (2017))
```

Task3 ran on local

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

```
rating = load '/home/acadgild/PIG/ratings.csv' using PigStorage(',') as (userid:int, movieid:int,  
rating:int, timestamp:int);
```

```
movierating = COGROUP movie by movieid, rating by movieid;
```

```
movierated = filter movierating by not isEmpty(rating);
```

```
movieratedtitle = foreach movierated generate flatten(movie.title);
```

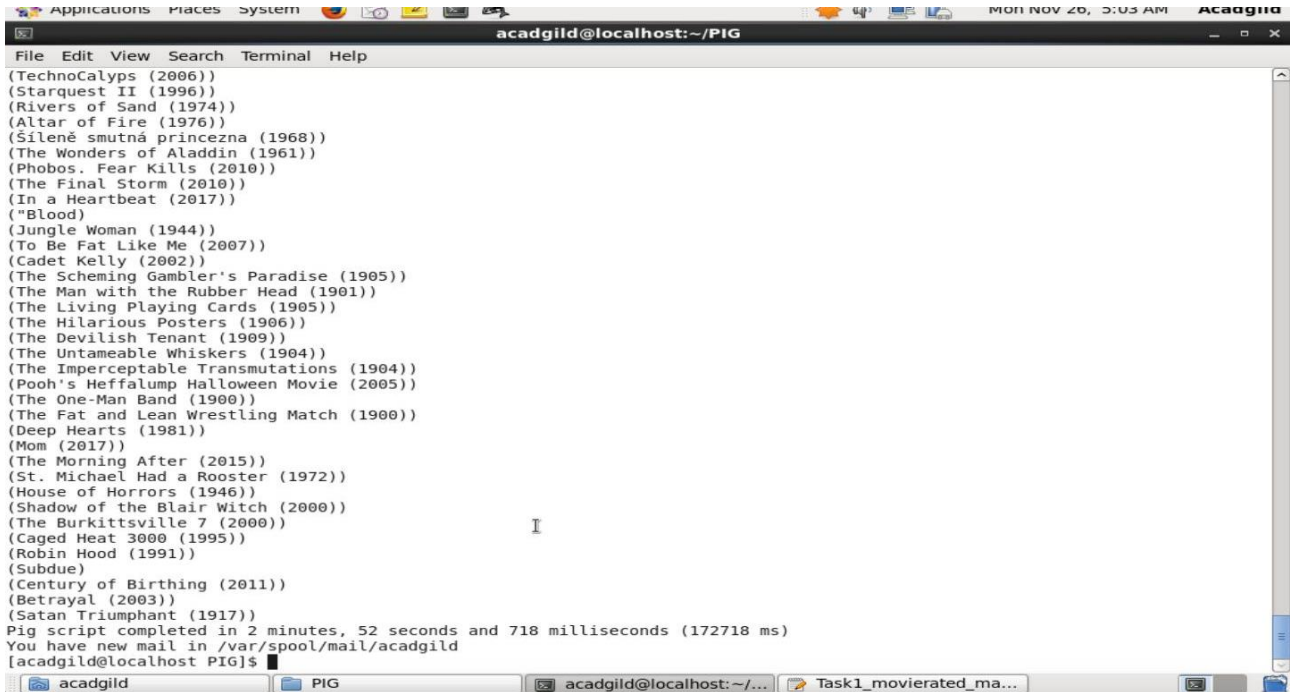
```
dump movieratedtitle;
```



The screenshot shows a terminal window titled 'acadgild@localhost: ~/PIG'. The output of the Pig script execution is displayed, including log messages from SLF4J, Pig version information (0.16.0), and warnings about deprecated features like 'fs.default.name' and 'mapred.job.tracker'. The script successfully executed the COGROUP, filter, and foreach operations.

```
File Edit View Search Terminal Help
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost PIG]$ pig -brief -x local Task1_movierated.pig
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/common/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/home/acadgild/install/hbase/hbase-1.2.6/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
18/11/26 05:04:43 INFO pig.ExecTypeProvider: Trying ExecType : LOCAL
18/11/26 05:04:43 INFO pig.ExecTypeProvider: Picked LOCAL as the ExecType
Apache Pig version 0.16.0 (r1746530) compiled Jun 01 2016, 23:10:49
Logging error messages to: /home/acadgild/PIG/pig_1543188883425.log
user.name is deprecated. Instead, use mapreduce.job.user.name
Default bootup file /home/acadgild/.pigbootup not found
mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
fs.default.name is deprecated. Instead, use fs.defaultFS
Connecting to hadoop file system at: file:///
Pig Script ID for the session: PIG-Task1_movierated.pig-89f2a69a-a2c1-49a4-9e6b-88900b17382f
ATS is disabled since yarn.timeline-service.enabled set to false
fs.default.name is deprecated. Instead, use fs.defaultFS
fs.default.name is deprecated. Instead, use fs.defaultFS
Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Pig features used in the script: COGROUP,FILTER
fs.default.name is deprecated. Instead, use fs.defaultFS
{RULES_ENABLED=[AddForEach, ColumnMapKeyPrune, ConstantCalculator, GroupByConstParallelSetter, LimitOptimizer, LoadTypeCastInserter, MergeFilter, MergeForEach, PartitionFilterOptimizer, PredicatePushdownOptimizer, PushDownForEachFlatten, PushDownFilter
```

## Output



```
acadgild@localhost:~/PIG
File Edit View Search Terminal Help
(TechnoCalyps (2006))
(Starquest II (1996))
(Rivers of Sand (1974))
(Altar of Fire (1976))
(Šíleně smutná princezna (1968))
(The Wonders of Aladdin (1961))
(Phobos. Fear Kills (2010))
(The Final Storm (2010))
(In a Heartbeat (2017))
("Blood)
(Jungle Woman (1944))
(To Be Fat Like Me (2007))
(Cadet Kelly (2002))
(The Scheming Gambler's Paradise (1905))
(The Man with the Rubber Head (1901))
(The Living Playing Cards (1905))
(The Hilarious Posters (1906))
(The Devilish Tenant (1909))
(The Untameable Whiskers (1904))
(The Imperceptable Transmutations (1904))
(Pooh's Heffalump Halloween Movie (2005))
(The One-Man Band (1900))
(The Fat and Lean Wrestling Match (1900))
(Deep Hearts (1981))
(Mom (2017))
(The Morning After (2015))
(St. Michael Had a Rooster (1972))
(House of Horrors (1946))
(Shadow of the Blair Witch (2000))
(The Burkittsville 7 (2000))
(Caged Heat 3000 (1995))
(Robin Hood (1991))
(Subdue)
(Century of Birthing (2011))
(Betrayal (2003))
(Satan Triumphant (1917))
Pig script completed in 2 minutes, 52 seconds and 718 milliseconds (172718 ms)
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost PIG]$
```

Task4 ran on map reduce

```
movie = load 'hdfs://localhost:8020/user/PIG/movies.csv' using PigStorage(',') as (movieid:int,
title:chararray, genres:chararray);
```

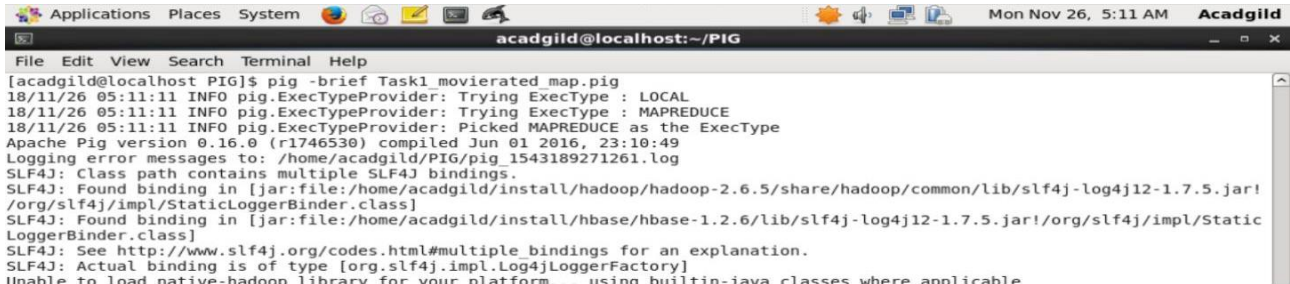
```
rating = load 'hdfs://localhost:8020/user/PIG/ratings.csv' using PigStorage(',') as (userid:int,
movieid:int, rating:int, timestamp:int);
```

```
movierating = COGROUP movie by movieid, rating by movieid;
```

```

movierated = filter movierating by not isEmpty(rating);
movieratedtitle = foreach movierated generate flatten(movie.title);
dump movieratedtitle;

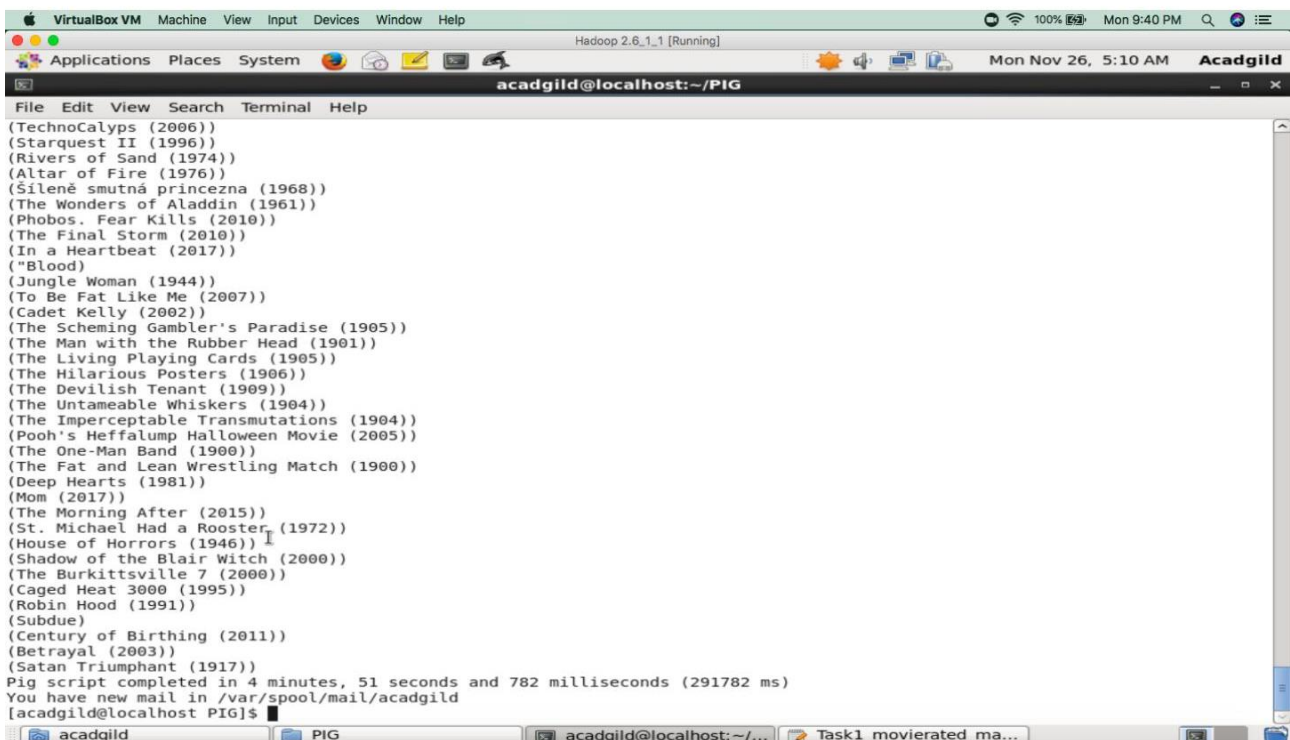
```



```

Applications  Places  System  Mon Nov 26, 5:11 AM  Acadgild
acadmild@localhost:~/PIG
File Edit View Search Terminal Help
[acadmild@localhost PIG]$ pig -brief Task1_movierated map.pig
18/11/26 05:11:11 INFO pig.ExecTypeProvider: Trying ExecType : LOCAL
18/11/26 05:11:11 INFO pig.ExecTypeProvider: Trying ExecType : MAPREDUCE
18/11/26 05:11:11 INFO pig.ExecTypeProvider: Picked MAPREDUCE as the ExecType
Apache Pig version 0.16.0 (r1746530) compiled Jun 01 2016, 23:10:49
Logging error messages to: /home/acadmild/PIG/pig_1543189271261.log
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/acadmild/install/hadoop/hadoop-2.6.5/share/hadoop/common/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/home/acadmild/install/hbase/hbase-1.2.6/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
Unable to load native-hadoop library for your platform - using builtin-java classes where applicable

```



```

VirtualBox VM  Machine  View  Input  Devices  Window  Help  Hadoop 2.6.1_1 [Running]  100%  Mon 9:40 PM
Applications  Places  System  Mon Nov 26, 5:10 AM  Acadgild
acadmild@localhost:~/PIG
File Edit View Search Terminal Help
(TechnoCalyps (2006))
(Starquest II (1996))
(Rivers of Sand (1974))
(Altar of Fire (1976))
(Šíleně smutná princezna (1968))
(The Wonders of Aladdin (1961))
(Phobos. Fear Kills (2010))
(The Final Storm (2010))
(In a Heartbeat (2017))
("Blood)
(Jungle Woman (1944))
(To Be Fat Like Me (2007))
(Cadet Kelly (2002))
(The Scheming Gambler's Paradise (1905))
(The Man with the Rubber Head (1901))
(The Living Playing Cards (1905))
(The Hilarious Posters (1906))
(The Devilish Tenant (1909))
(The Untameable Whiskers (1904))
(The Imperceptable Transmutations (1904))
(Pooh's Heffalump Halloween Movie (2005))
(The One-Man Band (1900))
(The Fat and Lean Wrestling Match (1900))
(Deep Hearts (1981))
(Mom (2017))
(The Morning After (2015))
(St. Michael Had a Rooster (1972))
(House of Horrors (1946))
(Shadow of the Blair Witch (2000))
(The Burkittsville 7 (2000))
(Caged Heat 3000 (1995))
(Robin Hood (1991))
(Subdue)
(Century of Birthing (2011))
(Betrayal (2003))
(Satan Triumphant (1917))
Pig script completed in 4 minutes, 51 seconds and 782 milliseconds (291782 ms)
You have new mail in /var/spool/mail/acadmild
[acadmild@localhost PIG]$

```



Task5 (How many times a movie was rated) ran on local

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

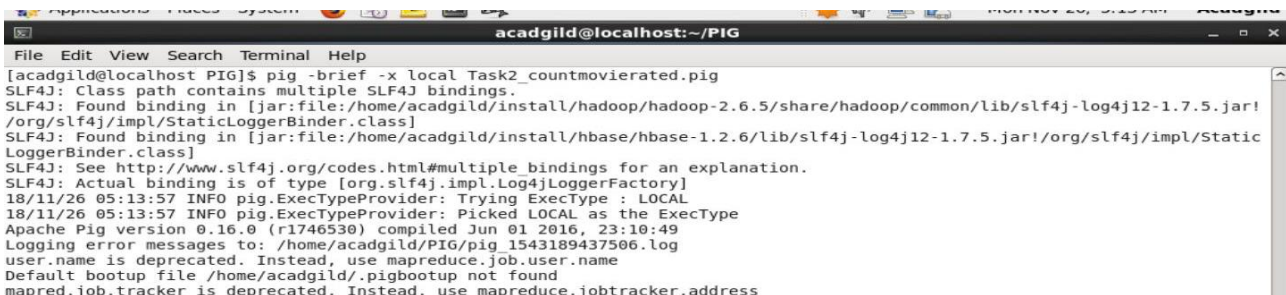
```
ratings = load '/home/acadgild/PIG/ratings.csv' using PigStorage(',') as (userid:int, movieid:int, rating:int, timestamp:int);
```

```
movierating = join movie by movieid left outer, ratings by movieid;
```

```
groupmovierating = group movierating by title;
```

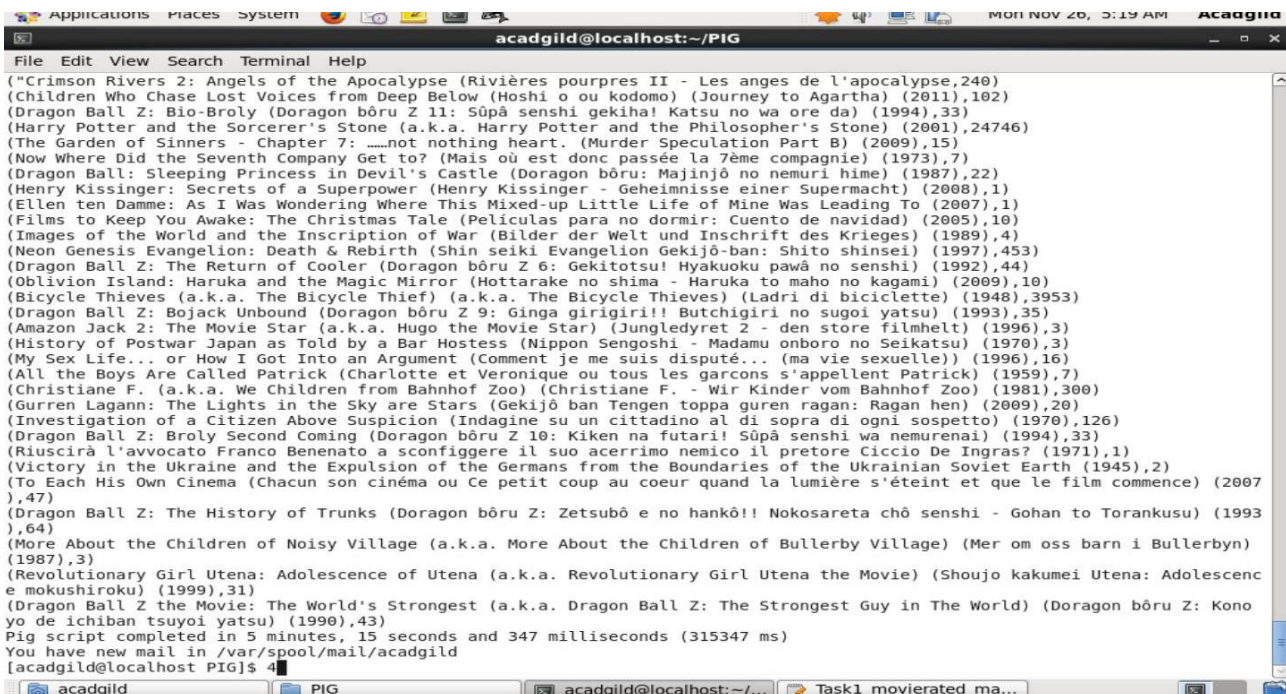
```
countmovierating = foreach groupmovierating generate group, COUNT(movierating.rating);
```

```
dump countmovierating;
```



```
acadgild@localhost:~/PIG
[acadgild@localhost PIG]$ pig -brief -x local Task2_countmovierated.pig
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/common/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/home/acadgild/install/hbase/hbase-1.2.6/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
18/11/26 05:13:57 INFO pig.ExecTypeProvider: Trying ExecType : LOCAL
18/11/26 05:13:57 INFO pig.ExecTypeProvider: Picked LOCAL as the ExecType
Apache Pig version 0.16.0 (r1746530) compiled Jun 01 2016, 23:10:49
Logging error messages to: /home/acadgild/PIG/pig_1543189437506.log
user.name is deprecated. Instead, use mapreduce.job.user.name
Default bootstrap file /home/acadgild/.pigbootstrap not found
mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
```

## Output



```
acadgild@localhost:~/PIG
("Crimson Rivers 2: Angels of the Apocalypse (Rivières pourpres II - Les anges de l'apocalypse,240)
(Children Who Chase Lost Voices from Deep Below (Hoshi o ou kodomo) (Journey to Agarth) (2011),102)
(Dragon Ball Z: Bio-Broly (Doragon bôru Z 11: Sûpâ senshi gekiha! Katsu no wa ore da) (1994),33)
(Harry Potter and the Sorcerer's Stone (a.k.a. Harry Potter and the Philosopher's Stone) (2001),24746)
(The Garden of Sinners - Chapter 7: ...not nothing heart. (Murder Speculation Part B) (2009),15)
(Now Where Did the Seventh Company Get to? (Mais où est donc passée la 7ème compagnie) (1973),7)
(Dragon Ball: Sleeping Princess in Devil's Castle (Doragon bôru: Majinjô no nemuri hime) (1987),22)
(Henry Kissinger: Secrets of a Superpower (Henry Kissinger - Geheimnisse einer Supermacht) (2008),1)
(Ellen ten Damme: As I Was Wondering Where This Mixed-up Little Life of Mine Was Leading To (2007),1)
(Films to Keep You Awake: The Christmas Tale (Películas para no dormir: Cuento de navidad) (2005),10)
(Images of the World and the Inscription of War (Bilder der Welt und Inschrift des Krieges) (1989),4)
(Neon Genesis Evangelion: Death & Rebirth (Shin seiki Evangelion Gekijô-ban: Shito shinsei) (1997),453)
(Dragon Ball Z: The Return of Cooler (Doragon bôru Z 6: Gekitotsu! Hyakuoku pawâ no senshi) (1992),44)
(Oblivion Island: Haruka and the Magic Mirror (Hottarake no shima - Haruka to maho no kagami) (2009),10)
(Bicycle Thieves (a.k.a. The Bicycle Thief) (a.k.a. The Bicycle Thieves) (Ladri di biciclette) (1948),3953)
(Dragon Ball Z: Bojack Unbound (Doragon bôru Z 9: Ginga girigiri!! Butchigiri no sugoi yatsu) (1993),35)
(Amazon Jack 2: The Movie Star (a.k.a. Hugo the Movie Star) (Jungledyret 2 - den store filmhelt) (1996),3)
(History of Postwar Japan as Told by a Bar Hostess (Nippon Sengoshi - Madamu onboro no Seikatsu) (1970),3)
(My Sex Life... or How I Got Into an Argument (Comment je me suis disputé... (ma vie sexuelle)) (1996),16)
(All the Boys Are Called Patrick (Charlotte et Veronique ou tous les garçons s'appellent Patrick) (1959),7)
(Christiane F. (a.k.a. We Children from Bahnhof Zoo) (Christiane F. - Wir Kinder vom Bahnhof Zoo) (1981),300)
(Gurren Lagann: The Lights in the Sky are Stars (Gekijô ban Tengen toppa guren ragan: Ragan hen) (2009),20)
(Investigation of a Citizen Above Suspicion (Indagine su un cittadino al di sopra di ogni sospetto) (1970),126)
(Dragon Ball Z: Broly Second Coming (Doragon bôru Z 10: Kiken na futari! Sûpâ senshi wa nemurenai) (1994),33)
(Riuscirà l'avvocato Franco Benenato a sconfiggere il suo acerrimo nemico il pretore Ciccio De Ingrassia? (1971),1)
(Victory in the Ukraine and the Expulsion of the Germans from the Boundaries of the Ukrainian Soviet Earth (1945),2)
(To Each His Own Cinema (Chacun son cinéma ou Ce petit coup au coeur quand la lumière s'éteint et que le film commence) (2007),47)
(Dragon Ball Z: The History of Trunks (Doragon bôru Z: Zetsubô e no hankô!! Nokosareta chô senshi - Gohan to Torankusu) (1993),64)
(More About the Children of Noisy Village (a.k.a. More About the Children of Bullerby Village) (Mer om oss barn i Bullerbyn) (1987),3)
(Revolutionary Girl Utena: Adolescence of Utena (a.k.a. Revolutionary Girl Utena the Movie) (Shoujo kakumei Utena: Adolescence e mokushiroku) (1999),31)
(Dragon Ball Z the Movie: The World's Strongest (a.k.a. Dragon Ball Z: The Strongest Guy in The World) (Doragon bôru Z: Kono yo de ichiban tsuyoi yatsu) (1990),43)
Pig script completed in 5 minutes, 15 seconds and 347 milliseconds (315347 ms)
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost PIG]$
```

## Task6

```
movie = load 'hdfs://localhost:8020/user/PIG/movies.csv' using PigStorage(',') as (movieid:int, title:chararray, genres:chararray);
```

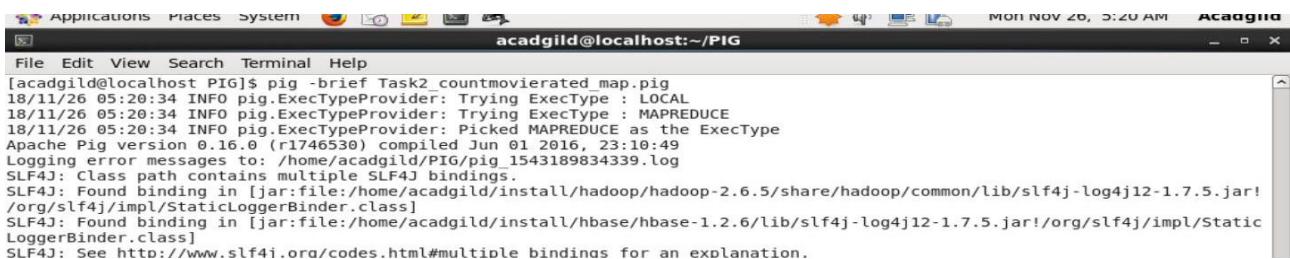
```
ratings = load 'hdfs://localhost:8020/user/PIG/ratings.csv' using PigStorage(',') as (userid:int, movieid:int, rating:int, timestamp:int);
```

```
movierating = join movie by movieid left outer, ratings by movieid;
```

```
groupmovierating = group movierating by title;
```

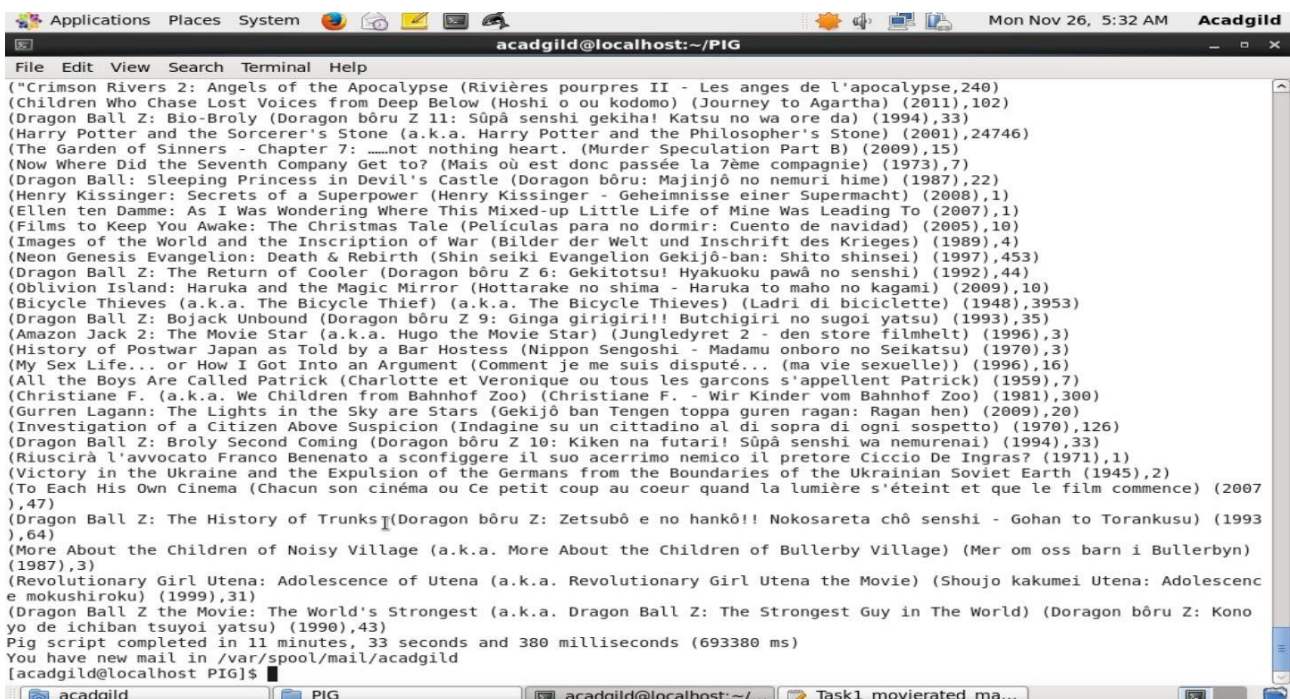
```
countmovierating = foreach groupmovierating generate group, COUNT(movierating.rating);
```

```
dump countmovierating;
```



```
acadgild@localhost:~/PIG
[acadgild@localhost PIG]$ pig -brief Task2_countmovierated_map.pig
18/11/26 05:20:34 INFO pig.ExecTypeProvider: Trying ExecType : LOCAL
18/11/26 05:20:34 INFO pig.ExecTypeProvider: Trying ExecType : MAPREDUCE
18/11/26 05:20:34 INFO pig.ExecTypeProvider: Picked MAPREDUCE as the ExecType
Apache Pig version 0.16.0 (r1746530) compiled Jun 01 2016, 23:10:49
Logging error messages to: /home/acadgild/PIG/pig_1543189834339.log
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/common/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/home/acadgild/install/hbase/hbase-1.2.6/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
```

## Output



```
acadgild@localhost:~/PIG
("Crimson Rivers 2: Angels of the Apocalypse (Rivières pourpres II - Les anges de l'apocalypse,240)
(Children Who Chase Lost Voices from Deep Below (Hoshi o ou kodomo) (Journey to Agartha) (2011),102)
(Dragon Ball Z: Bio-Broly (Doragon bôru Z 11: Sûpâ senshi gekiha! Katsu no wa ore da) (1994),33)
(Harry Potter and the Sorcerer's Stone (a.k.a. Harry Potter and the Philosopher's Stone) (2001),24746)
(The Garden of Sinners - Chapter 7: .....not nothing heart. (Murder Speculation Part B) (2009),15)
(Now Where Did the Seventh Company Get to? (Mais où est donc passée la 7ème compagnie) (1973),7)
(Dragon Ball: Sleeping Princess in Devil's Castle (Doragon bôru: Majinjô no nemuri hime) (1987),22)
(Henry Kissinger: Secrets of a Superpower (Henry Kissinger - Geheimnisse einer Supermacht) (2008),1)
(Ellen ten Damme: As I Was Wondering Where This Mixed-up Little Life of Mine Was Leading To (2007),1)
(Films to Keep You Awake: The Christmas Tale (Películas para no dormir: Cuento de navidad) (2005),10)
(Images of the World and the Inscription of War (Bilder der Welt und Inschrift des Krieges) (1989),4)
(Neon Genesis Evangelion: Death & Rebirth (Shin seiki Evangelion Gekijô-ban: Shito shinsei) (1997),453)
(Dragon Ball Z: The Return of Cooler (Doragon bôru Z 6: Gekitotsu! Hyakuoku pawâ no senshi) (1992),44)
(Oblivion Island: Haruka and the Magic Mirror (Hottarake no shima - Haruka to maho no kagami) (2009),10)
(Bicycle Thieves (a.k.a. The Bicycle Thief) (a.k.a. The Bicycle Thieves (Ladri di biciclette) (1948),3953)
(Dragon Ball Z: Bojack Unbound (Doragon bôru Z 9: Ginga girigiri!! Butchigiri no sugoi yatsu) (1993),35)
(Amazon Jack 2: The Movie Star (a.k.a. Hugo the Movie Star) (Jungledyret 2 - den store filmhelt) (1996),3)
(History of Postwar Japan as Told by a Bar Hostess (Nippon Sengoshi - Madamu onboro no Seikatsu) (1970),3)
(My Sex Life... or How I Got Into an Argument (Comment je me suis disputé... (ma vie sexuelle)) (1996),16)
(All the Boys Are Called Patrick (Charlotte et Veronique ou tous les garçons s'appellent Patrick) (1959),7)
(Christiane F. (a.k.a. We Children from Bahnhof Zoo) (Christiane F. - Wir Kinder vom Bahnhof Zoo) (1981),300)
(Gurrien Lagann: The Lights in the Sky are Stars (Gekijô ban Tengen toppa guren ragan: Ragan hen) (2009),20)
(Investigation of a Citizen Above Suspicion (Indagine su un cittadino al di sopra di ogni sospetto) (1970),126)
(Dragon Ball Z: Broly Second Coming (Doragon bôru Z 10: Kiken na futari! Sûpâ senshi wa nemurenai) (1994),33)
(Riuscirà l'avvocato Franco Benenato a sconfiggere il suo acerrimo nemico il pretore Ciccio De Ingrassia? (1971),1)
(Victory in the Ukraine and the Expulsion of the Germans from the Boundaries of the Ukrainian Soviet Earth (1945),2)
(To Each His Own Cinema (Chacun son cinéma ou Ce petit coup au coeur quand la lumière s'éteint et que le film commence) (2007),47)
(Dragon Ball Z: The History of Trunks (Doragon bôru Z: Zetsubô e no hankô!! Nokosareta chô senshi - Gohan to Torankusu) (1993),64)
(More About the Children of Noisy Village (a.k.a. More About the Children of Bullerby Village) (Mer om oss barn i Bullerbyn) (1987),3)
(Revolutionary Girl Utena: Adolescence of Utena (a.k.a. Revolutionary Girl Utena the Movie) (Shoujo kakumei Utena: Adolescence e mokushiroku) (1999),31)
(Dragon Ball Z the Movie: The World's Strongest (a.k.a. Dragon Ball Z: The Strongest Guy in The World) (Doragon bôru Z: Kono yo de ichiban tsuyoi yatsu) (1990),43)
Pig script completed in 11 minutes, 33 seconds and 380 milliseconds (693380 ms)
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost PIG]$
```



## Task7

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

ratings = load '/home/acadgild/PIG/ratings.csv' using PigStorage(',') as (userid:int, movieid:int, rating:int,
timestamp:int);

movierating = join movie by movieid, ratings by movieid;

groupmovierating = group movierating by title;

averagemovierating = foreach groupmovierating generate group, AVG(movierating.rating);

dump averagemovierating;
```



```
acacgild@localhost:~/PIG
You have new mail in /var/spool/mail/acacgild
[acacgild@localhost PIG]$ pig -brief -x local Task3_averagemovierating.pig
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/acacgild/install/hadoop/hadoop-2.6.5/share/hadoop/common/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/home/acacgild/install/hbase/hbase-1.2.6/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
18/11/26 05:37:42 INFO pig.ExecTypeProvider: Trying ExecType : LOCAL
18/11/26 05:37:42 INFO pig.ExecTypeProvider: Picked LOCAL as the ExecType
Apache Pig version 0.16.0 (r1746530) compiled Jun 01 2016, 23:10:49
```

## Output



```
acacgild@localhost:~/PIG
("Crimson Rivers 2: Angels of the Apocalypse (Rivières pourpres II - Les anges de l'apocalypse,2.8583333333333334)
(Children Who Chase Lost Voices from Deep Below (Hoshi o ou kodomo) (Journey to Agartha) (2011),3.2941176470588234)
(Dragon Ball Z: Bio-Broly (Doragon bôru Z 11: Sûpâ senshi gekiha! Katsu no wa ore da) (1994),2.787878787878788)
(Harry Potter and the Sorcerer's Stone (a.k.a. Harry Potter and the Philosopher's Stone) (2001),3.484603572294512)
(The Garden of Sinners - Chapter 7: .....not nothing heart. (Murder Speculation Part B) (2009),3.133333333333333)
(Now Where Did the Seventh Company Get to? (Mais où est donc passée la 7ème compagnie) (1973),3.0)
(Dragon Ball: Sleeping Princess in Devil's Castle (Doragon bôru Z: Majinjô no nemuri hime) (1987),2.5)
(Henry Kissinger: Secrets of a Superpower (Henry Kissinger - Geheimnisse einer Supermacht) (2008),3.0)
(Ellen ten Damme: As I Was Wondering Where This Mixed-up Little Life of Mine Was Leading To (2007),2.0)
(Films to Keep You Awake: The Christmas Tale (Películas para no dormir: Cuento de navidad) (2005),2.8)
(Images of the World and the Inscription of War (Bilder der Welt und Inschrift des Krieges) (1989),3.5)
(Neon Genesis Evangelion: Death & Rebirth (Shin seiki Evangelion Gekijô-ban: Shito shinsei) (1997),3.642384105960265)
(Dragon Ball Z: The Return of Cooler (Doragon bôru Z 6: Gekitotsu! Hyakuoku pawâ no senshi) (1992),3.0)
(Oblivion Island: Haruka and the Magic Mirror (Hottarake no shima - Haruka to maho no kagami) (2009),2.6)
(Bicycle Thieves (a.k.a. The Bicycle Thief) (a.k.a. The Bicycle Thieves) (Ladri di biciclette) (1948),3.987857323551733)
(Dragon Ball Z: Bojack Unbound (Doragon bôru Z 9: Ginga girigiri!! Butchigiri no sugoi yatsu) (1993),2.942857142857143)
(Amazon Jack 2: The Movie Star (a.k.a. Hugo the Movie Star) (Jungledyret 2 - den store filmhelt) (1996),2.6666666666666665)
(History of Postwar Japan as Told by a Bar Hostess (Nippon Sengoshi - Madamu onboro no Seikatsu) (1970),3.0)
(My Sex Life... or How I Got Into an Argument (Comment je me suis disputé... (ma vie sexuelle)) (1996),3.5)
(All the Boys Are Called Patrick (Charlotte et Veronique ou tous les garçons s'appellent Patrick) (1959),3.2857142857142856)
(Christiane F. (a.k.a. We Children from Bahnhof Zoo) (Christiane F. - Wir Kinder vom Bahnhof Zoo) (1981),3.4766666666666666)
(Gurren Lagann: The Lights in the Sky are Stars (Gekijô ban Tengen toppa guren ragan: Ragan hen) (2009),3.45)
(Investigation of a Citizen Above Suspicion (Indagine su un cittadino al di sopra di ogni sospetto) (1970),3.611111111111111)
(Dragon Ball Z: Broly Second Coming (Doragon bôru Z 10: Kiken na futari! Sûpâ senshi wa nemurenai) (1994),2.8484848484848486)
(Riuscirà l'avvocato Franco Benenato a sconfiggere il suo acerrimo nemico il pretore Ciccio De Ingrassia? (1971),3.0)
(Victory in the Ukraine and the Expulsion of the Germans from the Boundaries of the Ukrainian Soviet Earth (1945),1.5)
(To Each His Own Cinema (Chacun son cinéma ou Ce petit coup au coeur quand la lumière s'éteint et que le film commence) (2007),3.0638297872340425)
(Dragon Ball Z: The History of Trunks (Doragon bôru Z: Zetsubô e no hankô!! Nokosareta chô senshi - Gohan to Torankusu) (1993),3.390625)
(More About the Children of Noisy Village (a.k.a. More About the Children of Bullerby Village) (Mer om oss barn i Bullerbyn) (1987),3.333333333333333)
(Revolutionary Girl Utena: Adolescence of Utena (a.k.a. Revolutionary Girl Utena the Movie) (Shoujo kakumei Utena: Adolescenc e mukushiroku) (1999),3.4193548387096775)
(Dragon Ball Z: The Movie: The World's Strongest (a.k.a. Dragon Ball Z: The Strongest Guy in The World) (Doragon bôru Z: Kono yo de ichiban tsuyoi yatsu) (1990),2.7906976744186047)
Pig script completed in 5 minutes, 23 seconds and 478 milliseconds (323478 ms)
You have new mail in /var/spool/mail/acacgild
[acacgild@localhost PIG]$
```



## Task8

```
movie = load 'hdfs://localhost:8020/user/PIG/movies.csv' using PigStorage(',') as (movieid:int,
title:chararray, genres:chararray);

ratings = load 'hdfs://localhost:8020/user/PIG/ratings.csv' using PigStorage(',') as (userid:int, movieid:int,
rating:int, timestamp:int);

movierating = join movie by movieid, ratings by movieid;

groupmovierating = group movierating by title;

averagemovierating = foreach groupmovierating generate group, AVG(movierating.rating);

dump averagemovierating;
```

```
acadgild@localhost:~/PIG
File Edit View Search Terminal Help
[acadgild@localhost PIG]$ pig -brief Task3 averagemovierating_map.pig
18/11/26 05:44:29 INFO pig.ExecTypeProvider: Trying ExecType : LOCAL
18/11/26 05:44:29 INFO pig.ExecTypeProvider: Trying ExecType : MAPREDUCE
18/11/26 05:44:29 INFO pig.ExecTypeProvider: Picked MAPREDUCE as the ExecType
Apache Pig version 0.16.0 (r1746530) compiled Jun 01 2016, 23:10:49
Logging error messages to: /home/acadgild/PIG/pig_1543191269414.log
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/common/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/home/acadgild/install/hbase/hbase-1.2.6/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Default bootstrap file /home/acadgild/.pigbootstrap not found
mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
fs.default.name is deprecated. Instead, use fs.defaultFS
Connecting to hadoop file system at: hdfs://localhost:8020
```

## Output

```
acadgild@localhost:~/PIG
File Edit View Search Terminal Help
("Crimson Rivers 2: Angels of the Apocalypse (Rivières pourpres II - Les anges de l'apocalypse,2.8583333333333334)
(Children Who Chase Lost Voices from Deep Below (Hoshi o ou kodomo) (Journey to Agarth) (2011),3.2941176470588234)
(Dragon Ball Z: Bio-Broly (Doragon bôru Z 11: Sûpâ senshi gekiha! Katsu no wa ore da) (1994),2.787878787878788)
(Harry Potter and the Sorcerer's Stone (a.k.a. Harry Potter and the Philosopher's Stone) (2001),3.484603572294512)
(The Garden of Sinners - Chapter 7: .....not nothing heart. (Murder Speculation Part B) (2009),3.133333333333333)
(Now Where Did the Seventh Company Get to? (Mais où est donc passée la 7ème compagnie) (1973),3.0)
(Dragon Ball: Sleeping Princess in Devil's Castle (Doragon bôru: Majinjô no nemuri hime) (1987),2.5)
(Henry Kissinger: Secrets of a Superpower (Henry Kissinger - Geheimnisse einer Supermacht) (2008),3.0)
(Ellen ten Damme: As I Was Wondering Where This Mixed-up Little Life of Mine Was Leading To (2007),2.0)
(Films to Keep You Awake: The Christmas Tale (Películas para no dormir: Cuento de navidad) (2005),2.8)
(Images of the World and the Inscription of War (Bilder der Welt und Inschrift des Krieges) (1989),3.5)
(Neon Genesis Evangelion: Death & Rebirth (Shin seiki Evangelion Gekijô-ban: Shito shinsei) (1997),3.642384105960265)
(Dragon Ball Z: The Return of Cooler (Doragon bôru Z 6: Gekitotsu! Hyakuoku pawâ no senshi) (1992),3.0)
(Oblivion Island: Haruka and the Magic Mirror (Hottarake no shima - Haruka to maho no kagami) (2009),2.6)
(Bicycle Thieves (a.k.a. The Bicycle Thief) (a.k.a. The Bicycle Thieves) (Ladri di biciclette) (1948),3.987857223551733)
(Dragon Ball Z: Bojack Unbound (Doragon bôru Z 9: Ginga girigiri!! Butchigiri no sugoi yatsu) (1993),2.942857142857143)
(Amazon Jack 2: The Movie Star (a.k.a. Hugo the Movie Star) (Jungledyret 2 - den store filmhelt) (1996),2.666666666666666)
(History of Postwar Japan as Told by a Bar Hostess (Nippon Sengoshi - Madamu onbôro no Seikatsu) (1970),3.0)
(My Sex Life... or How I Got Into an Argument (Comment je me suis disputé... (ma vie sexuelle)) (1996),3.5)
(All the Boys Are Called Patrick (Charlotte et Veronique ou tous les garçons s'appellent Patrick) (1959),3.2857142857142856)
(Christiane F. (a.k.a. We Children from Bahnhof Zoo) (Christiane F. - Wir Kinder vom Bahnhof Zoo) (1981),3.476666666666666)
(Gurren Lagann: The Lights in the Sky are Stars (Gekijô ban Tengen toppa guren ragan: Ragan hen) (2009),3.45)
(Investigation of a Citizen Above Suspicion (Indagine su un cittadino al di sopra di ogni sospetto) (1970),3.611111111111111)
(Dragon Ball Z: Broly Second Coming (Doragon bôru Z 10: Kiken na futari! Sûpâ senshi wa nemurenai) (1994),2.848484848484848)
(Riuscirà l'avvocato Franco Benenato a sconfiggere il suo acerrimo nemico il pretore Ciccio De Ingrass? (1971),3.0)
(Victory in the Ukraine and the Expulsion of the Germans from the Boundaries of the Ukrainian Soviet Earth (1945),1.5)
(To Each His Own Cinema (Chacun son cinéma ou Ce petit coup au coeur quand la lumière s'éteint et que le film commence) (2007
```