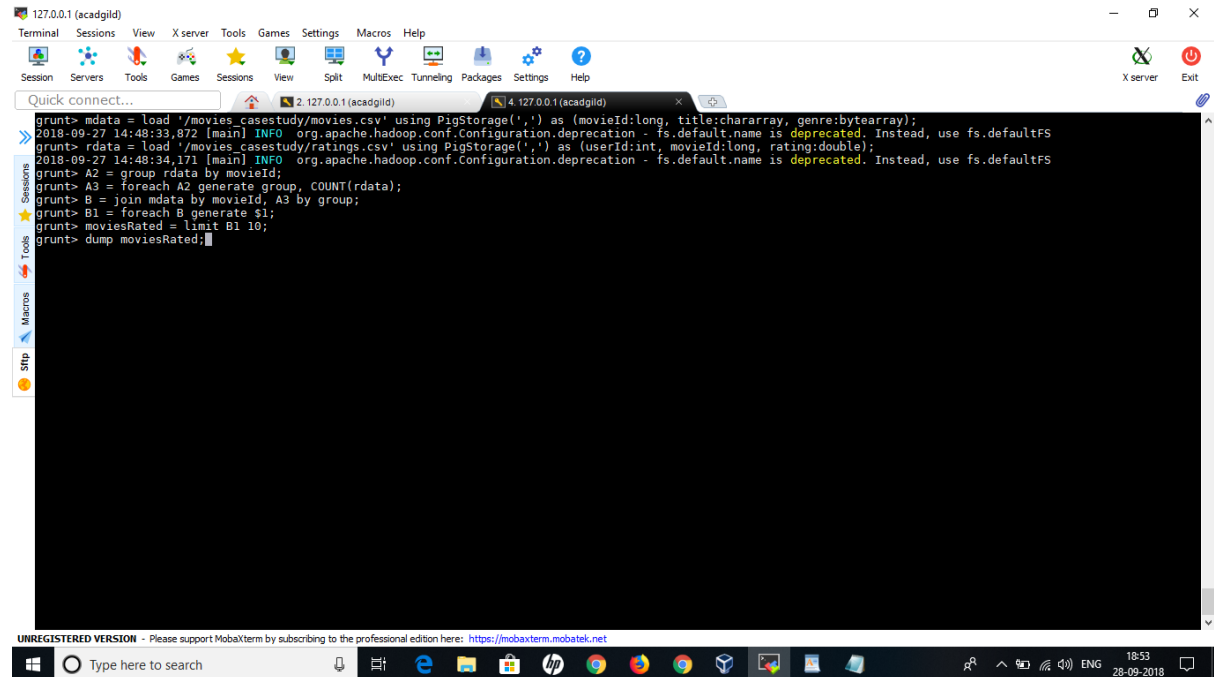


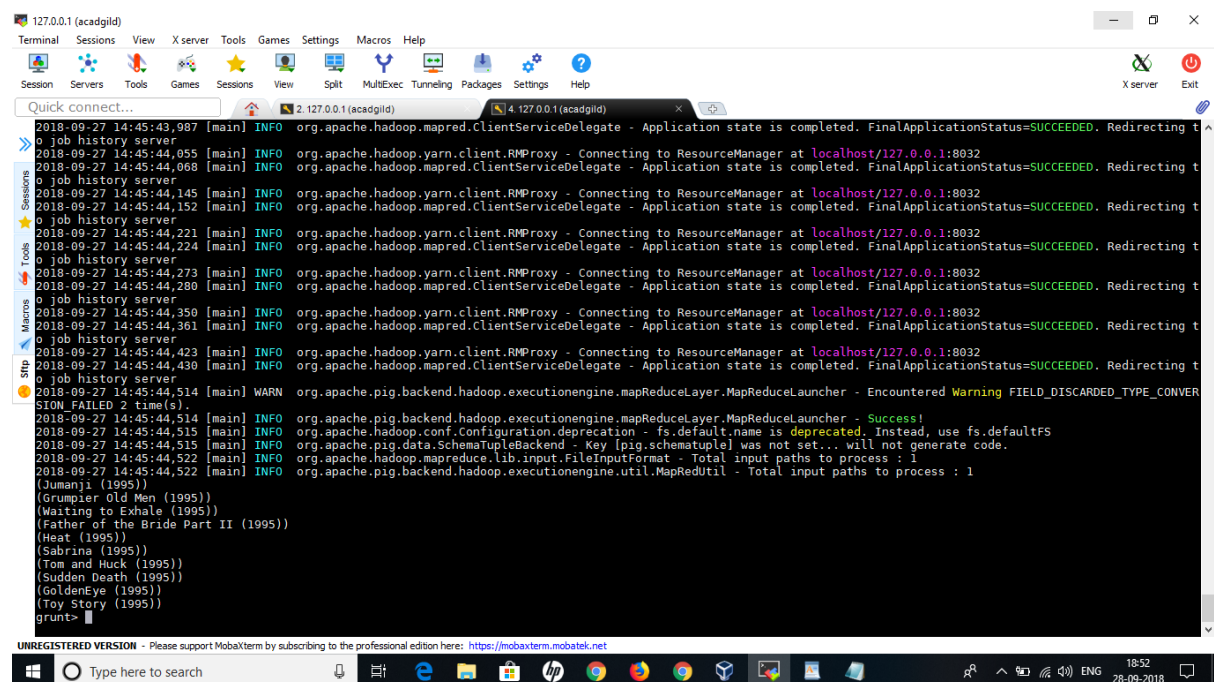
Case Study Movie Rating.

Task 1: What are the movie titles that the user has rated?

Solution: Please see the below screenshots:



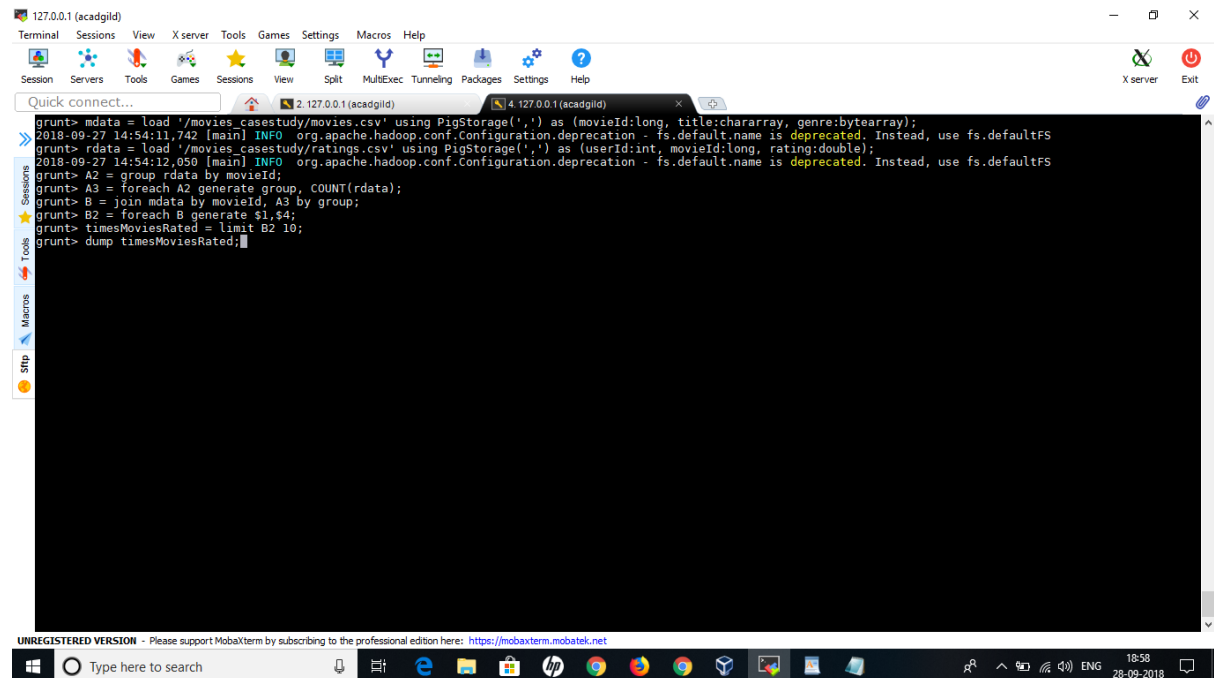
```
127.0.0.1 (acadgild)
Terminal Sessions View X server Tools Games Settings Macros Help
Quick connect...
2. 127.0.0.1 (acadgild) 4. 127.0.0.1 (acadgild)
grunt> mdata = load '/movies_casestudy/movies.csv' using PigStorage(',') as (movieId:long, title:chararray, genre:bytearray);
2018-09-27 14:48:33,872 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
grunt> rdata = load '/movies_casestudy/ratings.csv' using PigStorage(',') as (userId:int, movieId:long, rating:double);
2018-09-27 14:48:34,171 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
grunt> A2 = group rdata by movieId;
grunt> A3 = foreach A2 generate group, COUNT(rdata);
grunt> B = join mdata by movieId, A3 by group;
grunt> B1 = foreach B generate $1;
grunt> moviesRated = limit B1 10;
grunt> dump moviesRated;
```



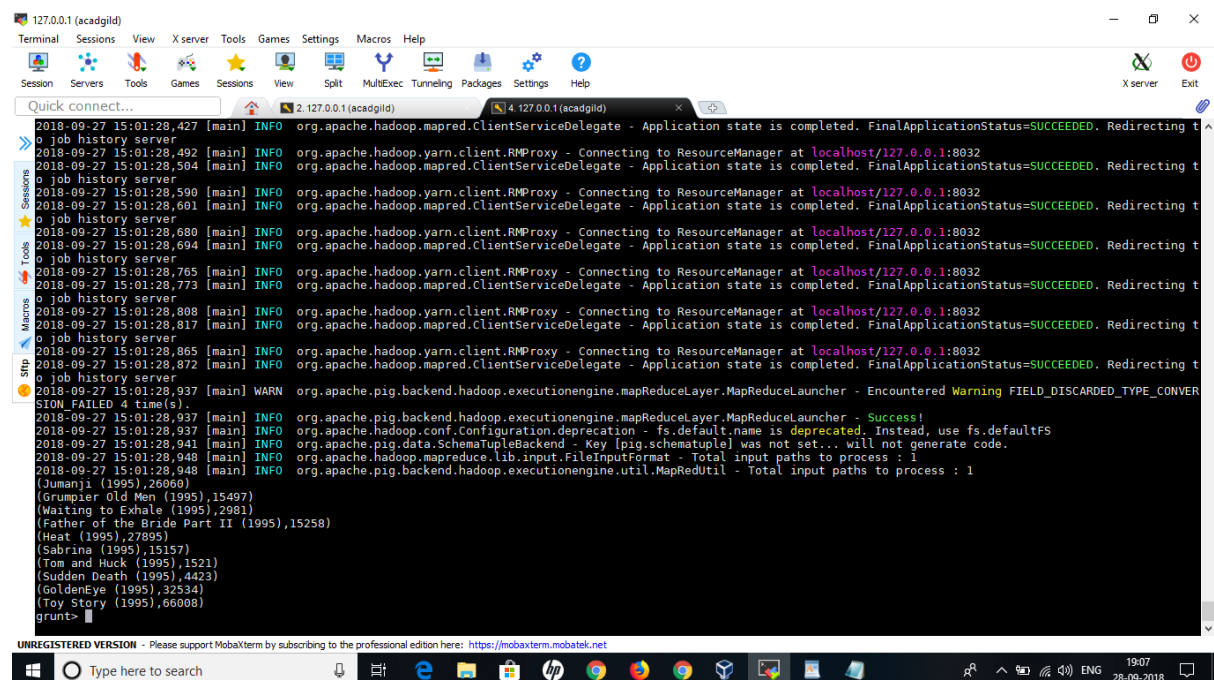
```
127.0.0.1 (acadgild)
Terminal Sessions View X server Tools Games Settings Macros Help
Quick connect...
2. 127.0.0.1 (acadgild) 4. 127.0.0.1 (acadgild)
2018-09-27 14:45:43,987 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting t
o job history server
2018-09-27 14:45:44,055 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-09-27 14:45:44,068 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting t
o job history server
2018-09-27 14:45:44,145 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-09-27 14:45:44,152 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting t
o job history server
2018-09-27 14:45:44,221 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-09-27 14:45:44,224 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting t
o job history server
2018-09-27 14:45:44,273 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-09-27 14:45:44,280 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting t
o job history server
2018-09-27 14:45:44,350 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-09-27 14:45:44,361 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting t
o job history server
2018-09-27 14:45:44,423 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-09-27 14:45:44,430 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting t
o job history server
2018-09-27 14:45:44,514 [main] WARN org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Encountered Warning FIELD_DISCARDED_TYPE_CONVER
SION FAILED 2 time(s).
2018-09-27 14:45:44,514 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2018-09-27 14:45:44,515 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2018-09-27 14:45:44,515 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.
2018-09-27 14:45:44,522 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2018-09-27 14:45:44,522 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(Jumanji (1995))
(Grumpier Old Men (1995))
(Waiting to Exhale (1995))
(Father of the Bride Part II (1995))
(Heat (1995))
(Sabrina (1995))
(Tom and Huck (1995))
(Sudden Death (1995))
(GoldenEye (1995))
(Toy Story (1995))
grunt>
```

Task 2: How many times a movie has been rated by the user?

Solution: Please see the below screenshot:



```
127.0.0.1 (acadgild)
Terminal Sessions View X server Tools Games Settings Macros Help
Session Servers Tools Games Sessions View Split MultiExec Tunneling Packages Settings Help
Quick connect...
2. 127.0.0.1 (acadgild) 4. 127.0.0.1 (acadgild)
> grunt> mdata = load '/movies/casestudy/movies.csv' using PigStorage(',') as (movieId:long, title:chararray, genre:bytearray);
2018-09-27 14:54:11,742 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
> grunt> rdata = load '/movies/casestudy/ratings.csv' using PigStorage(',') as (userId:int, movieId:long, rating:double);
2018-09-27 14:54:12,050 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
> grunt> A2 = group rdata by movieId;
> grunt> A3 = foreach A2 generate group, COUNT(rdata);
> grunt> B = join mdata by movieId, A3 by group;
> grunt> B2 = foreach B generate $1,$4;
> grunt> timesMoviesRated = limit B2 10;
> grunt> dump timesMoviesRated;
```



```
127.0.0.1 (acadgild)
Terminal Sessions View X server Tools Games Settings Macros Help
Session Servers Tools Games Sessions View Split MultiExec Tunneling Packages Settings Help
Quick connect...
2. 127.0.0.1 (acadgild) 4. 127.0.0.1 (acadgild)
2018-09-27 15:01:28,427 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting t
o job history server
2018-09-27 15:01:28,492 [main] INFO org.apache.hadoop.yarn.client.RMPProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-09-27 15:01:28,504 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting t
o job history server
2018-09-27 15:01:28,590 [main] INFO org.apache.hadoop.yarn.client.RMPProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-09-27 15:01:28,601 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting t
o job history server
2018-09-27 15:01:28,680 [main] INFO org.apache.hadoop.yarn.client.RMPProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-09-27 15:01:28,694 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting t
o job history server
2018-09-27 15:01:28,765 [main] INFO org.apache.hadoop.yarn.client.RMPProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-09-27 15:01:28,773 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting t
o job history server
2018-09-27 15:01:28,808 [main] INFO org.apache.hadoop.yarn.client.RMPProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-09-27 15:01:28,817 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting t
o job history server
2018-09-27 15:01:28,865 [main] INFO org.apache.hadoop.yarn.client.RMPProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-09-27 15:01:28,872 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting t
o job history server
2018-09-27 15:01:28,937 [main] WARN org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Encountered Warning FIELD_DISCARDED_TYPE_CONVER
SION_FAILED 4 time(s).
2018-09-27 15:01:28,937 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2018-09-27 15:01:28,937 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2018-09-27 15:01:28,941 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schema.tuple] was not set... will not generate code.
2018-09-27 15:01:28,948 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2018-09-27 15:01:28,948 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(Jumanji (1995),26060)
(Grumpier Old Men (1995),15497)
(Waiting to Exhale (1995),2981)
(Father of the Bride Part II (1995),15258)
(Heat (1995),27895)
(Sabrina (1995),15157)
(Tom and Huck (1995),1521)
(Sudden Death (1995),4423)
(GoldenEye (1995),32534)
(Toy Story (1995),66008)
> grunt>
```

Task 3: What is the average rating given for a movie?

Solution: Please see the below screenshots:

127.0.0.1 (acadgild)

Terminal Sessions View X server Tools Games Settings Macros Help

Session Servers Tools Games Sessions View Split MultiExec Tunneling Packages Settings Help

Quick connect...

```
grunt> mdata = load '/movies_casestudy/movies.csv' using PigStorage(',') as (movieId:long, title:chararray, genre:bytearray);
2018-09-27 15:11:02,155 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
grunt> rdata = load '/movies_casestudy/ratings.csv' using PigStorage(',') as (userId:int, movieId:long, rating:double);
2018-09-27 15:11:02,545 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
grunt> A2 = group rdata by movieId;
grunt> A3 = foreach A2 generate group, rdata.rating;
grunt> A3 = foreach A2 generate group, AVG(rdata.rating);
grunt> B = join mdata by movieId, A3 by group;
grunt> B3 = foreach B generate $1,$4;
grunt> ratingAvg = limit B3 10;
grunt> dump ratingAvg;
```

UNREGISTERED VERSION - Please support MobaXterm by subscribing to the professional edition here: <https://mobaxterm.mobatek.net>

Type here to search

127.0.0.1 (acadgild)

Terminal Sessions View X server Tools Games Settings Macros Help

Session Servers Tools Games Sessions View Split MultiExec Tunneling Packages Settings Help

Quick connect...

```
2018-09-27 15:18:56,723 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting t
o job history server
2018-09-27 15:18:56,795 [main] INFO org.apache.hadoop.yarn.client.RMPProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-09-27 15:18:56,809 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting t
o job history server
2018-09-27 15:18:56,901 [main] INFO org.apache.hadoop.yarn.client.RMPProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-09-27 15:18:56,912 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting t
o job history server
2018-09-27 15:18:56,979 [main] INFO org.apache.hadoop.yarn.client.RMPProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-09-27 15:18:56,986 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting t
o job history server
2018-09-27 15:18:57,054 [main] INFO org.apache.hadoop.yarn.client.RMPProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-09-27 15:18:57,060 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting t
o job history server
2018-09-27 15:18:57,102 [main] INFO org.apache.hadoop.yarn.client.RMPProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-09-27 15:18:57,111 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting t
o job history server
2018-09-27 15:18:57,146 [main] INFO org.apache.hadoop.yarn.client.RMPProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-09-27 15:18:57,155 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting t
o job history server
2018-09-27 15:18:57,198 [main] WARN org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Encountered Warning FIELD_DISCARDED_TYPE_CONVER
SION_FAILED 4 time(s).
2018-09-27 15:18:57,198 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2018-09-27 15:18:57,199 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2018-09-27 15:18:57,199 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.
2018-09-27 15:18:57,203 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2018-09-27 15:18:57,203 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(Jumanji (1995),3.2369531849577897)
(Grumpier Old Men (1995),3.1755501064722202)
(Waiting to Exhale (1995),2.8757128480375713)
(Father of the Bride Part II (1995),3.079564818455892)
(Heat (1995),3.841763756945689)
(Sabrina (1995),3.372108297082167)
(Tom and Huck (1995),3.1281913214990137)
(Sudden Death (1995),3.008365362875876)
(GoldenEye (1995),3.431840536054589)
(Toy Story (1995),3.8881574960610834)
grunt>
```

UNREGISTERED VERSION - Please support MobaXterm by subscribing to the professional edition here: <https://mobaxterm.mobatek.net>

Type here to search