

# 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?

## Dataset:

## SourceCode:

Syntax for command: CaseStudyIUseCase2Driver <input path1> <input path2> <output path>

```
[acadgild@localhost Documents]$ hadoop fs -ls /hadoopdata/hdfs/casestudy
18/02/07 16:19:47 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 6 items
-rw-r--r-- 1 acadgild supergroup 344861061 2018-02-07 16:18 /hadoopdata/hdfs/casestudy/genome-scores.csv
-rw-r--r-- 1 acadgild supergroup 18103 2018-02-07 16:18 /hadoopdata/hdfs/casestudy/genome-tags.csv
-rw-r--r-- 1 acadgild supergroup 989107 2018-02-07 16:18 /hadoopdata/hdfs/casestudy/links.csv
-rw-r--r-- 1 acadgild supergroup 2283410 2018-02-07 16:18 /hadoopdata/hdfs/casestudy/movies.csv
-rw-r--r-- 1 acadgild supergroup 709550327 2018-02-07 16:18 /hadoopdata/hdfs/casestudy/ratings.csv
-rw-r--r-- 1 acadgild supergroup 27113729 2018-02-07 16:18 /hadoopdata/hdfs/casestudy/tags.csv
[acadgild@localhost Documents]$ hadoop jar session7.jar com.acadgild.hadoop.session7Todo.CaseStudyIUseCase2Driver /hadoopdata/hdfs/casestudy/movies.csv /hadoopdata/hdfs/casestudy/ratings.csv /hadoopdata/hdfs/out
18/02/07 16:37:12 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/02/07 16:37:15 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/02/07 16:37:17 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
18/02/07 16:37:19 INFO input.FileInputFormat: Total input paths to process : 1
18/02/07 16:37:19 INFO input.FileInputFormat: Total input paths to process : 1
18/02/07 16:37:19 INFO mapreduce.JobSubmitter: number of splits:7
18/02/07 16:37:23 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1517997096477_0001
18/02/07 16:37:24 INFO impl.YarnClientImpl: Submitted application application_1517997096477_0001
18/02/07 16:37:25 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1517997096477_0001/
18/02/07 16:37:25 INFO mapreduce.Job: Running job: job_1517997096477_0001
18/02/07 16:37:42 INFO mapreduce.Job: Job job_1517997096477_0001 running in uber mode : false
18/02/07 16:37:42 INFO mapreduce.Job: map 0% reduce 0%
18/02/07 16:38:38 INFO mapreduce.Job: map 4% reduce 0%
18/02/07 16:38:46 INFO mapreduce.Job: map 16% reduce 0%
18/02/07 16:38:48 INFO mapreduce.Job: map 21% reduce 0%
18/02/07 16:38:52 INFO mapreduce.Job: map 26% reduce 0%
18/02/07 16:38:54 INFO mapreduce.Job: map 28% reduce 0%
18/02/07 16:38:55 INFO mapreduce.Job: map 29% reduce 0%
18/02/07 16:38:56 INFO mapreduce.Job: map 30% reduce 0%
18/02/07 16:38:57 INFO mapreduce.Job: map 32% reduce 0%
18/02/07 16:39:01 INFO mapreduce.Job: map 33% reduce 0%
18/02/07 16:39:05 INFO mapreduce.Job: map 34% reduce 0%
18/02/07 16:39:10 INFO mapreduce.Job: map 39% reduce 0%
18/02/07 16:39:22 INFO mapreduce.Job: map 40% reduce 0%
18/02/07 16:39:27 INFO mapreduce.Job: map 42% reduce 0%
18/02/07 16:39:30 INFO mapreduce.Job: map 43% reduce 0%
18/02/07 16:39:33 INFO mapreduce.Job: map 46% reduce 0%
18/02/07 16:39:37 INFO mapreduce.Job: map 49% reduce 0%
18/02/07 16:39:39 INFO mapreduce.Job: map 54% reduce 0%
18/02/07 16:39:43 INFO mapreduce.Job: map 56% reduce 0%
18/02/07 16:39:46 INFO mapreduce.Job: map 58% reduce 0%
18/02/07 16:39:50 INFO mapreduce.Job: map 59% reduce 0%
18/02/07 16:39:52 INFO mapreduce.Job: map 60% reduce 0%
18/02/07 16:39:53 INFO mapreduce.Job: map 61% reduce 0%
```

# ACADGILD – SESSION7(EXPLORING APACHE PIG)

```

18/02/07 16:39:53 INFO mapreduce.Job: map 61% reduce 0%
18/02/07 16:39:56 INFO mapreduce.Job: map 62% reduce 0%
18/02/07 16:40:06 INFO mapreduce.Job: map 63% reduce 0%
18/02/07 16:40:07 INFO mapreduce.Job: map 64% reduce 0%
18/02/07 16:40:10 INFO mapreduce.Job: map 65% reduce 0%
18/02/07 16:40:13 INFO mapreduce.Job: map 67% reduce 0%
18/02/07 16:40:16 INFO mapreduce.Job: map 70% reduce 0%
18/02/07 16:40:18 INFO mapreduce.Job: map 71% reduce 0%
18/02/07 16:40:19 INFO mapreduce.Job: map 72% reduce 0%
18/02/07 16:40:22 INFO mapreduce.Job: map 73% reduce 0%
18/02/07 16:40:23 INFO mapreduce.Job: map 75% reduce 0%
18/02/07 16:40:25 INFO mapreduce.Job: map 76% reduce 0%
18/02/07 16:40:26 INFO mapreduce.Job: map 78% reduce 0%
18/02/07 16:40:28 INFO mapreduce.Job: map 79% reduce 0%
18/02/07 16:40:29 INFO mapreduce.Job: map 80% reduce 0%
18/02/07 16:40:33 INFO mapreduce.Job: map 82% reduce 0%
18/02/07 16:40:33 INFO mapreduce.Job: map 83% reduce 0%
18/02/07 16:40:37 INFO mapreduce.Job: map 85% reduce 0%
18/02/07 16:40:40 INFO mapreduce.Job: map 86% reduce 0%
18/02/07 16:40:47 INFO mapreduce.Job: map 100% reduce 0%
18/02/07 16:40:56 INFO mapreduce.Job: map 100% reduce 24%
18/02/07 16:40:59 INFO mapreduce.Job: map 100% reduce 67%
18/02/07 16:41:14 INFO mapreduce.Job: map 100% reduce 68%
18/02/07 16:41:23 INFO mapreduce.Job: map 100% reduce 69%
18/02/07 16:41:35 INFO mapreduce.Job: map 100% reduce 70%
18/02/07 16:41:46 INFO mapreduce.Job: map 100% reduce 71%
18/02/07 16:41:56 INFO mapreduce.Job: map 100% reduce 72%
18/02/07 16:42:08 INFO mapreduce.Job: map 100% reduce 73%
18/02/07 16:42:17 INFO mapreduce.Job: map 100% reduce 74%
18/02/07 16:42:29 INFO mapreduce.Job: map 100% reduce 75%
18/02/07 16:42:38 INFO mapreduce.Job: map 100% reduce 76%
18/02/07 16:42:50 INFO mapreduce.Job: map 100% reduce 77%
18/02/07 16:42:59 INFO mapreduce.Job: map 100% reduce 78%
18/02/07 16:43:11 INFO mapreduce.Job: map 100% reduce 79%
18/02/07 16:43:20 INFO mapreduce.Job: map 100% reduce 80%
18/02/07 16:43:32 INFO mapreduce.Job: map 100% reduce 81%
18/02/07 16:43:44 INFO mapreduce.Job: map 100% reduce 82%
18/02/07 16:43:56 INFO mapreduce.Job: map 100% reduce 83%
18/02/07 16:44:10 INFO mapreduce.Job: map 100% reduce 84%
18/02/07 16:44:19 INFO mapreduce.Job: map 100% reduce 85%
18/02/07 16:44:31 INFO mapreduce.Job: map 100% reduce 86%
18/02/07 16:44:40 INFO mapreduce.Job: map 100% reduce 87%
18/02/07 16:44:52 INFO mapreduce.Job: map 100% reduce 88%
18/02/07 16:45:04 INFO mapreduce.Job: map 100% reduce 89%
18/02/07 16:45:13 INFO mapreduce.Job: map 100% reduce 90%
18/02/07 16:45:22 INFO mapreduce.Job: map 100% reduce 91%

18/02/07 16:45:34 INFO mapreduce.Job: map 100% reduce 92%
18/02/07 16:45:43 INFO mapreduce.Job: map 100% reduce 93%
18/02/07 16:45:55 INFO mapreduce.Job: map 100% reduce 94%
18/02/07 16:46:04 INFO mapreduce.Job: map 100% reduce 95%
18/02/07 16:46:16 INFO mapreduce.Job: map 100% reduce 96%
18/02/07 16:46:25 INFO mapreduce.Job: map 100% reduce 97%
18/02/07 16:46:37 INFO mapreduce.Job: map 100% reduce 98%
18/02/07 16:46:46 INFO mapreduce.Job: map 100% reduce 99%
18/02/07 16:46:58 INFO mapreduce.Job: map 100% reduce 100%
18/02/07 16:47:03 INFO mapreduce.Job: Job job_1517997096477_0001 completed successfully
18/02/07 16:47:03 INFO mapreduce.Job: Counter: 50
File System Counters
  FILE: Number of bytes read=961694264
  FILE: Number of bytes written=1457487790
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=711856301
  HDFS: Number of bytes written=1669001
  HDFS: Number of read operations=24
  HDFS: Number of large read operations=0
  HDFS: Number of write operations=2
Job Counters
  Killed map tasks=2
  Launched map tasks=8
  Launched reduce tasks=1
  Data-local map tasks=8
  Total time spent by all maps in occupied slots (ms)=1031593
  Total time spent by all reduces in occupied slots (ms)=386350
  Total time spent by all map tasks (ms)=1031593
  Total time spent by all reduce tasks (ms)=386350
  Total vcore-milliseconds taken by all map tasks=1031593
  Total vcore-milliseconds taken by all reduce tasks=386350
  Total megabyte-milliseconds taken by all map tasks=1056351232
  Total megabyte-milliseconds taken by all reduce tasks=395622400
Map-Reduce Framework
  Map input records=26070134
  Map output records=26070132
  Map output bytes=442789828
  Map output materialized bytes=494930141
  Input split bytes=2084
  Combine input records=0
  Combine output records=0
  Reduce input groups=45843
  Reduce shuffle bytes=494930141
  Reduce input records=26070132
  Reduce output records=45843
  Spilled Records=76775523
  Shuffled Maps =7
  Failed Shuffles=0
  Merged Map outputs=7
  GC time elapsed (ms)=13449
  CPU time spent (ms)=442200
  Physical memory (bytes) snapshot=1461657600
  Virtual memory (bytes) snapshot=16445816832
  Total committed heap usage (bytes)=1032732672
Shuffle Errors
  BAD_ID=0
  CONNECTION=0
  IO_ERROR=0
  WRONG_LENGTH=0
  WRONG_MAP=0
  WRONG_REDUCE=0
File Input Format Counters
  Bytes Read=0
File Output Format Counters
  Bytes Written=1669001
[acadgild@localhost Documents]$

```

## OUTPUT:

```
acadgild@localhost:~/Documents
[acadgild@localhost Documents]$ hadoop fs -ls /hadoopdata/hdfs/out
18/02/07 16:55:35 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 2 items
-rw-r--r-- 1 acadgild supergroup 0 2018-02-07 16:47 /hadoopdata/hdfs/out/_SUCCESS
-rw-r--r-- 1 acadgild supergroup 1669001 2018-02-07 16:46 /hadoopdata/hdfs/out/part-r-00000
[acadgild@localhost Documents]$ hadoop fs -cat /hadoopdata/hdfs/out/part-r-00000 | head
18/02/07 16:55:47 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     3.232304
Curdled (1996)       217      3.099078
*Comic 1             4.000000
Op 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 Documents]$
```