

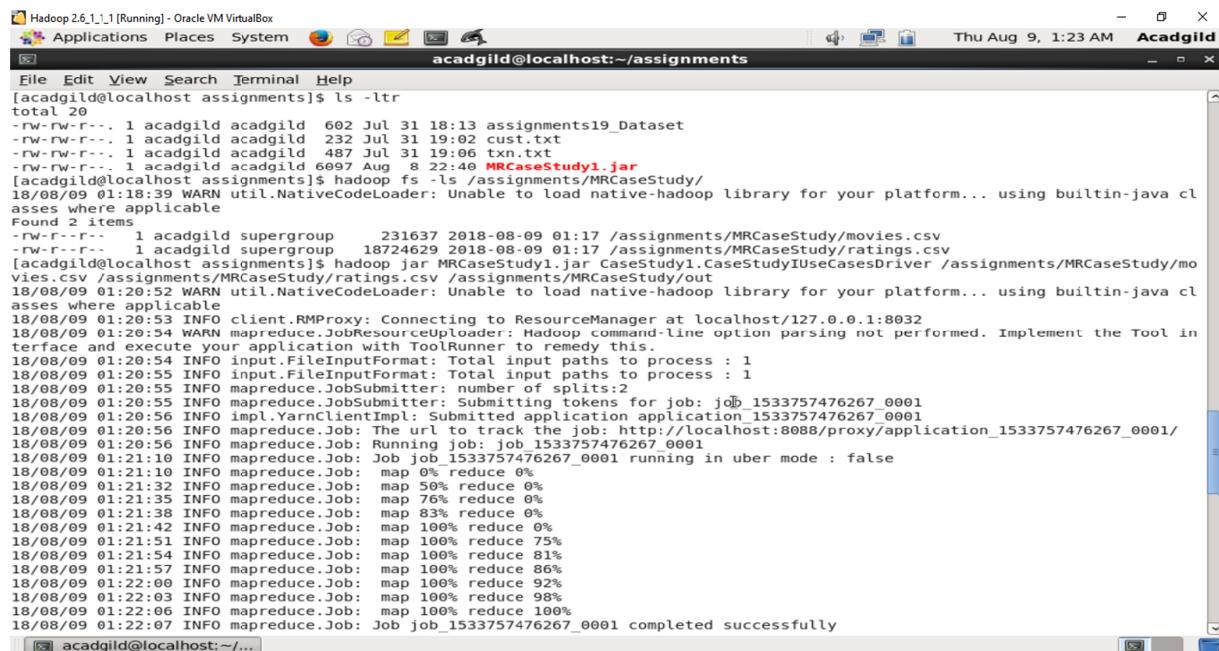
Problem Statement:

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?

Used an extract of the data from the given dataset to perform this task since the given dataset was getting hung in VM.



```
[acadgild@localhost assignments]$ ls -ltr
total 20
-rw-r--r-- 1 acadgild acadgild 602 Jul 31 18:13 assignments19_Dataset
-rw-r--r-- 1 acadgild acadgild 232 Jul 31 19:02 cust.txt
-rw-r--r-- 1 acadgild acadgild 487 Jul 31 19:06 txn.txt
-rw-r--r-- 1 acadgild acadgild 6097 Aug  8 22:40 MRCaseStudy1.jar
[acadgild@localhost assignments]$ hadoop fs -ls /assignments/MRCaseStudy/
18/08/09 01:18:39 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 231637 2018-08-09 01:17 /assignments/MRCaseStudy/movies.csv
-rw-r--r-- 1 acadgild supergroup 18724629 2018-08-09 01:17 /assignments/MRCaseStudy/ratings.csv
[acadgild@localhost assignments]$ hadoop jar MRCaseStudy1.jar CaseStudy1.CaseStudyIUseCasesDriver /assignments/MRCaseStudy/movies.csv /assignments/MRCaseStudy/ratings.csv /assignments/MRCaseStudy/out
18/08/09 01:20:52 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/08/09 01:20:53 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/08/09 01:20:54 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
18/08/09 01:20:54 INFO input.FileInputFormat: Total input paths to process : 1
18/08/09 01:20:55 INFO input.FileInputFormat: Total input paths to process : 1
18/08/09 01:20:55 INFO mapreduce.JobSubmitter: number of splits:2
18/08/09 01:20:55 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1533757476267_0001
18/08/09 01:20:56 INFO impl.YarnClientImpl: Submitted application application_1533757476267_0001
18/08/09 01:20:56 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1533757476267_0001/
18/08/09 01:20:56 INFO mapreduce.Job: Running job: job_1533757476267_0001
18/08/09 01:21:10 INFO mapreduce.Job: Job job_1533757476267_0001 running in uber mode : false
18/08/09 01:21:10 INFO mapreduce.Job: map 0% reduce 0%
18/08/09 01:21:32 INFO mapreduce.Job: map 50% reduce 0%
18/08/09 01:21:35 INFO mapreduce.Job: map 76% reduce 0%
18/08/09 01:21:38 INFO mapreduce.Job: map 83% reduce 0%
18/08/09 01:21:42 INFO mapreduce.Job: map 100% reduce 0%
18/08/09 01:21:51 INFO mapreduce.Job: map 100% reduce 75%
18/08/09 01:21:54 INFO mapreduce.Job: map 100% reduce 81%
18/08/09 01:21:57 INFO mapreduce.Job: map 100% reduce 86%
18/08/09 01:22:00 INFO mapreduce.Job: map 100% reduce 92%
18/08/09 01:22:03 INFO mapreduce.Job: map 100% reduce 98%
18/08/09 01:22:06 INFO mapreduce.Job: map 100% reduce 100%
18/08/09 01:22:07 INFO mapreduce.Job: Job job_1533757476267_0001 completed successfully
```

```
Hadoop 2.6.1_1 [Running] - Oracle VM VirtualBox  
Applications Places System acadgild@localhost:~/assignments  
File Edit View Search Terminal Help  
18/08/09 01:22:07 INFO mapreduce.Job: Job job_1533757476267_0001 completed successfully  
18/08/09 01:22:08 INFO mapreduce.Job: Counters: 50  
File System Counters  
FILE: Number of bytes read=12519442  
FILE: Number of bytes written=25362415  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=18956812  
HDFS: Number of bytes written=168958  
HDFS: Number of read operations=9  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
Job Counters  
Killed map tasks=1  
Launched map tasks=3  
Launched reduce tasks=1  
Data-local map tasks=3  
Total time spent by all maps in occupied slots (ms)=54782  
Total time spent by all reduces in occupied slots (ms)=30169  
Total time spent by all map tasks (ms)=54782  
Total time spent by all reduce tasks (ms)=30169  
Total vcore-milliseconds taken by all map tasks=54782  
Total vcore-milliseconds taken by all reduce tasks=30169  
Total megabyte-milliseconds taken by all map tasks=56096768  
Total megabyte-milliseconds taken by all reduce tasks=30893056  
Map-Reduce Framework  
Map input records=1053483  
Map output records=729054  
Map output bytes=11061328  
Map output materialized bytes=12519448  
Input split bytes=546  
Combine input records=0  
Combine output records=0  
Reduce input groups=4906  
Reduce shuffle bytes=12519448  
Reduce input records=729054  
Reduce output records=4906  
Spilled Records=1458108  
Acadgild@localhost:~/...
```

```
Spilled Records=1458108  
Shuffled Maps =2  
Failed Shuffles=0  
Merged Map outputs=2  
GC time elapsed (ms)=1815  
CPU time spent (ms)=33180  
Physical memory (bytes) snapshot=641720320  
Virtual memory (bytes) snapshot=6250831872  
Total committed heap usage (bytes)=462422016  
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=168958  
You have new mail in /var/spool/mail/acadgild  
[acadgild@localhost assignments]$ █  
Acadgild@localhost:~/...
```

```
[acadgild@localhost assignments]$ hadoop fs -cat /assignments/MRCaseStudy/out/part-r-00000 | head -20
18/08/09 01:25:46 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
Toy Story (1995)      2681    3.897799
GoldenEye (1995)       1300    3.428077
City Hall (1996)       204     3.235294
Curled (1996)        14      3.250000
"Associate            3      2.000000
Ed's Next Move (1996) 7      3.428571
Extreme Measures (1996) 72     3.027778
"Glimmer Man          62     2.588710
D3: The Mighty Ducks (1996)   90     2.472222
"Chamber              76     3.309211
"Apple Dumpling Gang  71     2.992958
"Davy Crockett         36     3.361111
Escape to Witch Mountain (1975) 104    3.163462
Bottle Rocket (1996)     184    3.788043
"Love Bug              95     3.005263
Herbie Rides Again (1974)    68     2.801471
Old Yeller (1957)        151    3.483444
"Parent Trap           210    3.350000
Pollyanna (1960)         62     3.427419
Homeward Bound: The Incredible Journey (1993)  171    3.266082
cat: Unable to write to output stream.
[acadgild@localhost assignments]$
[acadgild@localhost assignments]$
```

acadgild@localhost:~/...