

CASESTUDY1 ASSIGNMENT

Problem Statement:

What are the movie titles that the user has rated?

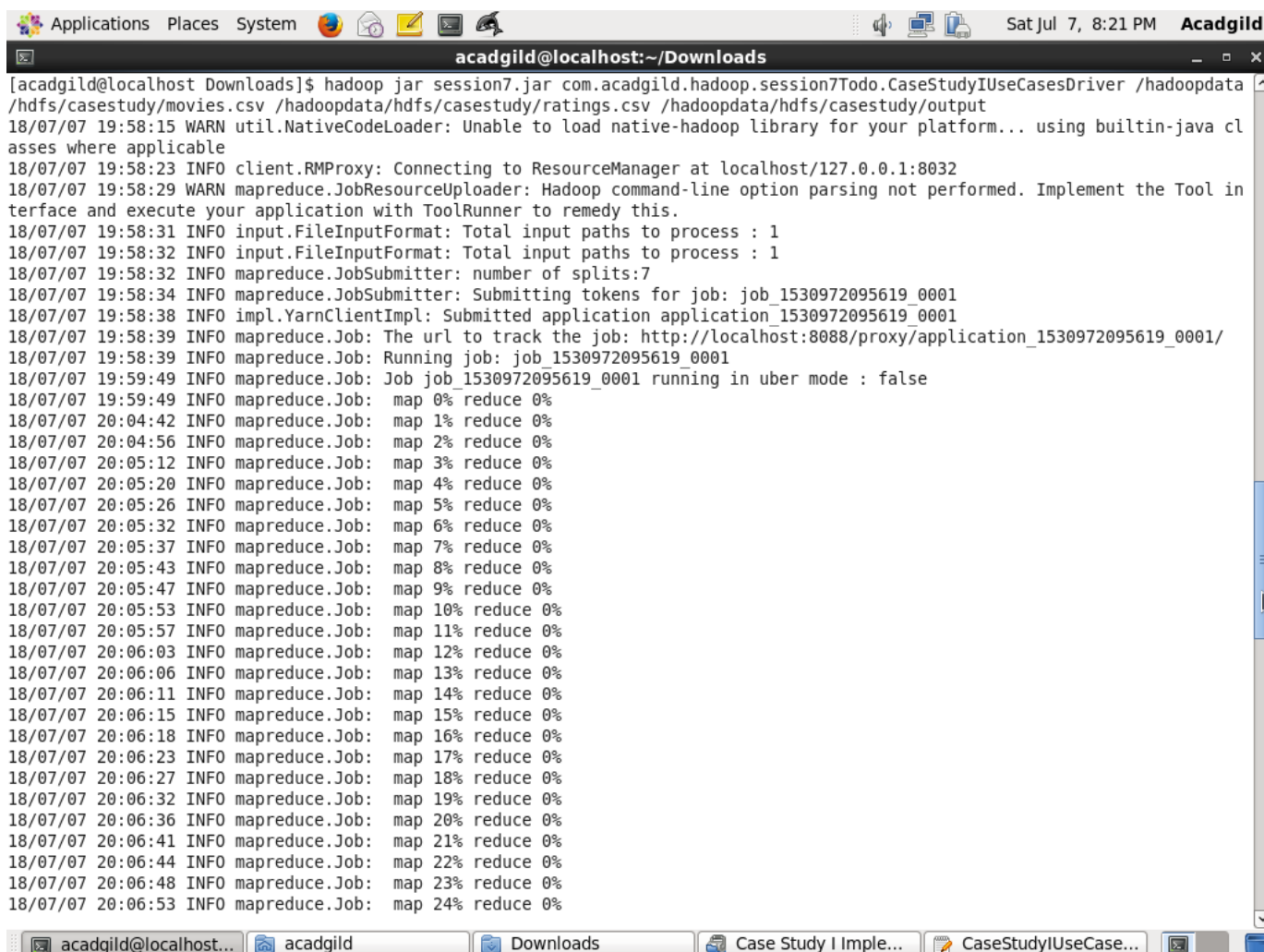
How many times a movie has been rated by the user?

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

```
18/07/07 19:45:14 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
Found 6 items
-rw-r--r-- 1 acadgild supergroup 344861061 2018-07-07 19:41 /hadoopdata/hdfs/casestudy/genome-scores.csv
-rw-r--r-- 1 acadgild supergroup 18103 2018-07-07 19:41 /hadoopdata/hdfs/casestudy/genome-tags.csv
-rw-r--r-- 1 acadgild supergroup 989107 2018-07-07 19:41 /hadoopdata/hdfs/casestudy/links.csv
-rw-r--r-- 1 acadgild supergroup 2283410 2018-07-07 19:41 /hadoopdata/hdfs/casestudy/movies.csv
-rw-r--r-- 1 acadgild supergroup 709550327 2018-07-07 19:43 /hadoopdata/hdfs/casestudy/ratings.csv
-rw-r--r-- 1 acadgild supergroup 27113729 2018-07-07 19:43 /hadoopdata/hdfs/casestudy/tags.csv
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost Downloads]$ cd ..
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$ hadoop fs -mkdir /hadoopdata/hdfs/casestudyoutput
18/07/07 19:52:29 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$ hadoop fs -mkdir /hadoopdata/hdfs/casestudy/output
18/07/07 19:53:10 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
[acadgild@localhost ~]$ hadoop jar session7.jar com.acadgild.hadoop.session7Todo.caseStudyUserCasesDriver /hadoopdata/hdfs/ca
sestudy/movies.csv /hadoopdata/hdfs/casestudy/ratings.csv /hadoopdata/hdfs/casestudy/output
Not a valid JAR: /home/acadgild/session7.jar
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$
```

acadgild@localhost:~

CASESTUDY1 ASSIGNMENT



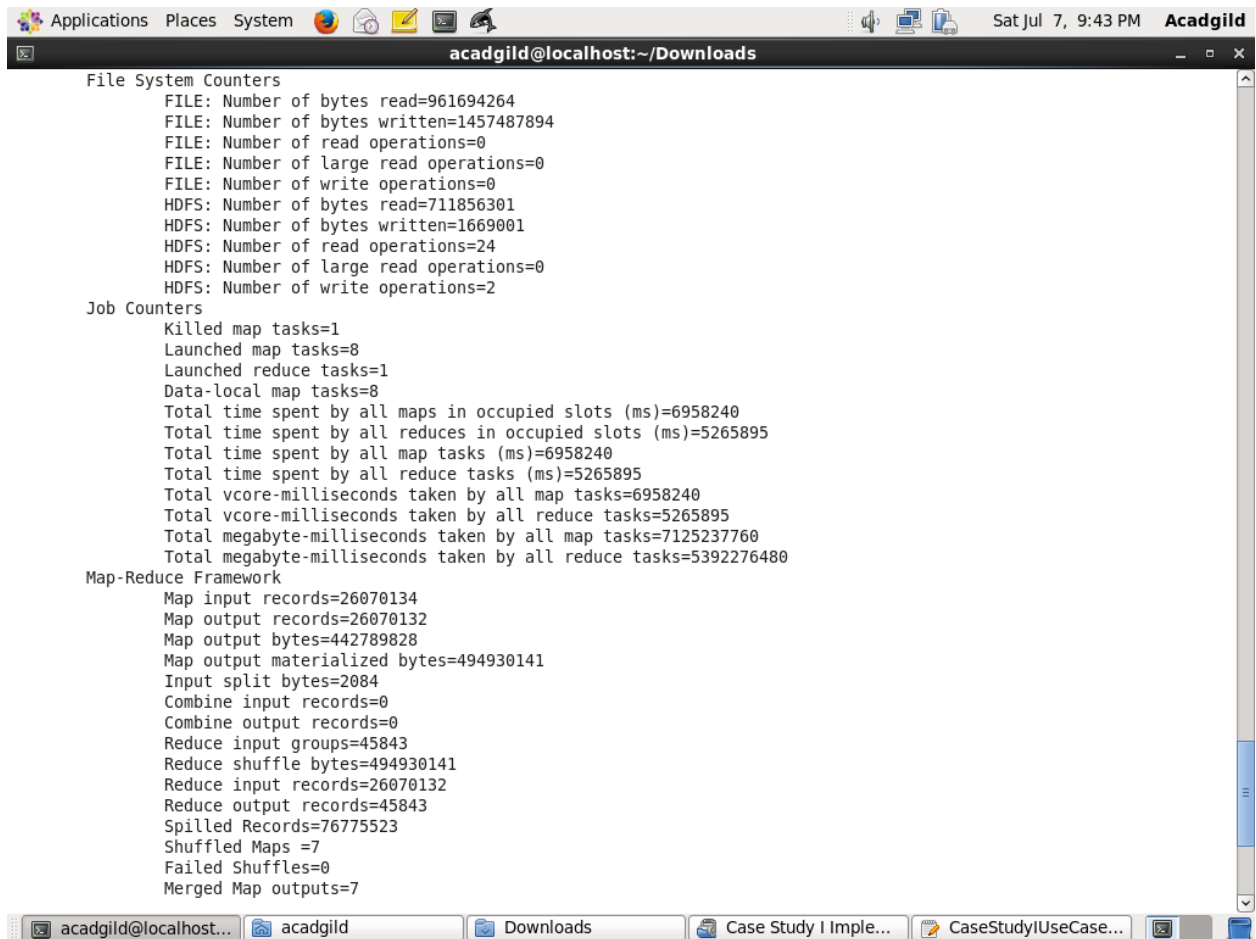
The screenshot shows a Linux desktop environment with a terminal window open. The terminal title is 'acadgild@localhost:~/Downloads'. The command executed is 'hadoop jar session7.jar com.acadgild.hadoop.session7Todo.CaseStudyIUseCasesDriver /hadoopdata/hdfs/casestudy/movies.csv /hadoopdata/hdfs/casestudy/ratings.csv /hadoopdata/hdfs/casestudy/output'. The logs show the job submission and progress, including warnings about native code loading and information about the job configuration and progress.

```
[acadgild@localhost Downloads]$ hadoop jar session7.jar com.acadgild.hadoop.session7Todo.CaseStudyIUseCasesDriver /hadoopdata/hdfs/casestudy/movies.csv /hadoopdata/hdfs/casestudy/ratings.csv /hadoopdata/hdfs/casestudy/output
18/07/07 19:58:15 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/07/07 19:58:23 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/07/07 19:58:29 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
18/07/07 19:58:31 INFO input.FileInputFormat: Total input paths to process : 1
18/07/07 19:58:32 INFO input.FileInputFormat: Total input paths to process : 1
18/07/07 19:58:32 INFO mapreduce.JobSubmitter: number of splits:7
18/07/07 19:58:34 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1530972095619_0001
18/07/07 19:58:38 INFO impl.YarnClientImpl: Submitted application application_1530972095619_0001
18/07/07 19:58:39 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1530972095619_0001/
18/07/07 19:58:39 INFO mapreduce.Job: Running job: job_1530972095619_0001
18/07/07 19:59:49 INFO mapreduce.Job: Job job_1530972095619_0001 running in uber mode : false
18/07/07 19:59:49 INFO mapreduce.Job: map 0% reduce 0%
18/07/07 20:04:42 INFO mapreduce.Job: map 1% reduce 0%
18/07/07 20:04:56 INFO mapreduce.Job: map 2% reduce 0%
18/07/07 20:05:12 INFO mapreduce.Job: map 3% reduce 0%
18/07/07 20:05:20 INFO mapreduce.Job: map 4% reduce 0%
18/07/07 20:05:26 INFO mapreduce.Job: map 5% reduce 0%
18/07/07 20:05:32 INFO mapreduce.Job: map 6% reduce 0%
18/07/07 20:05:37 INFO mapreduce.Job: map 7% reduce 0%
18/07/07 20:05:43 INFO mapreduce.Job: map 8% reduce 0%
18/07/07 20:05:47 INFO mapreduce.Job: map 9% reduce 0%
18/07/07 20:05:53 INFO mapreduce.Job: map 10% reduce 0%
18/07/07 20:05:57 INFO mapreduce.Job: map 11% reduce 0%
18/07/07 20:06:03 INFO mapreduce.Job: map 12% reduce 0%
18/07/07 20:06:06 INFO mapreduce.Job: map 13% reduce 0%
18/07/07 20:06:11 INFO mapreduce.Job: map 14% reduce 0%
18/07/07 20:06:15 INFO mapreduce.Job: map 15% reduce 0%
18/07/07 20:06:18 INFO mapreduce.Job: map 16% reduce 0%
18/07/07 20:06:23 INFO mapreduce.Job: map 17% reduce 0%
18/07/07 20:06:27 INFO mapreduce.Job: map 18% reduce 0%
18/07/07 20:06:32 INFO mapreduce.Job: map 19% reduce 0%
18/07/07 20:06:36 INFO mapreduce.Job: map 20% reduce 0%
18/07/07 20:06:41 INFO mapreduce.Job: map 21% reduce 0%
18/07/07 20:06:44 INFO mapreduce.Job: map 22% reduce 0%
18/07/07 20:06:48 INFO mapreduce.Job: map 23% reduce 0%
18/07/07 20:06:53 INFO mapreduce.Job: map 24% reduce 0%
```

CASESTUDY1 ASSIGNMENT

```
Applications Places System Sat Jul 7, 9:43 PM Acadgild
acadgild@localhost:~/Downloads
18/07/07 20:21:23 INFO mapreduce.Job: map 100% reduce 24%
18/07/07 20:21:31 INFO mapreduce.Job: map 100% reduce 29%
18/07/07 20:21:49 INFO mapreduce.Job: map 100% reduce 36%
18/07/07 20:21:54 INFO mapreduce.Job: map 100% reduce 41%
18/07/07 20:21:57 INFO mapreduce.Job: map 100% reduce 55%
18/07/07 20:22:00 INFO mapreduce.Job: map 100% reduce 67%
18/07/07 20:24:17 INFO mapreduce.Job: map 100% reduce 68%
18/07/07 20:26:51 INFO mapreduce.Job: map 100% reduce 69%
18/07/07 20:29:15 INFO mapreduce.Job: map 100% reduce 70%
18/07/07 20:31:43 INFO mapreduce.Job: map 100% reduce 71%
18/07/07 20:34:27 INFO mapreduce.Job: map 100% reduce 72%
18/07/07 20:36:48 INFO mapreduce.Job: map 100% reduce 73%
18/07/07 20:39:09 INFO mapreduce.Job: map 100% reduce 74%
18/07/07 20:41:29 INFO mapreduce.Job: map 100% reduce 75%
18/07/07 20:43:50 INFO mapreduce.Job: map 100% reduce 76%
18/07/07 20:46:07 INFO mapreduce.Job: map 100% reduce 77%
18/07/07 20:48:31 INFO mapreduce.Job: map 100% reduce 78%
18/07/07 20:50:57 INFO mapreduce.Job: map 100% reduce 79%
18/07/07 20:53:14 INFO mapreduce.Job: map 100% reduce 80%
18/07/07 20:55:28 INFO mapreduce.Job: map 100% reduce 81%
18/07/07 20:58:00 INFO mapreduce.Job: map 100% reduce 82%
18/07/07 21:00:20 INFO mapreduce.Job: map 100% reduce 83%
18/07/07 21:02:54 INFO mapreduce.Job: map 100% reduce 84%
18/07/07 21:05:17 INFO mapreduce.Job: map 100% reduce 85%
18/07/07 21:07:44 INFO mapreduce.Job: map 100% reduce 86%
18/07/07 21:10:07 INFO mapreduce.Job: map 100% reduce 87%
18/07/07 21:12:24 INFO mapreduce.Job: map 100% reduce 88%
18/07/07 21:14:58 INFO mapreduce.Job: map 100% reduce 89%
18/07/07 21:17:29 INFO mapreduce.Job: map 100% reduce 90%
18/07/07 21:19:52 INFO mapreduce.Job: map 100% reduce 91%
18/07/07 21:22:12 INFO mapreduce.Job: map 100% reduce 92%
18/07/07 21:24:36 INFO mapreduce.Job: map 100% reduce 93%
18/07/07 21:26:53 INFO mapreduce.Job: map 100% reduce 94%
18/07/07 21:29:24 INFO mapreduce.Job: map 100% reduce 95%
18/07/07 21:31:49 INFO mapreduce.Job: map 100% reduce 96%
18/07/07 21:34:12 INFO mapreduce.Job: map 100% reduce 97%
18/07/07 21:36:30 INFO mapreduce.Job: map 100% reduce 98%
18/07/07 21:38:57 INFO mapreduce.Job: map 100% reduce 99%
18/07/07 21:41:24 INFO mapreduce.Job: map 100% reduce 100%
18/07/07 21:42:36 INFO mapreduce.Job: Job job_1530972095619_0001 completed successfully
```

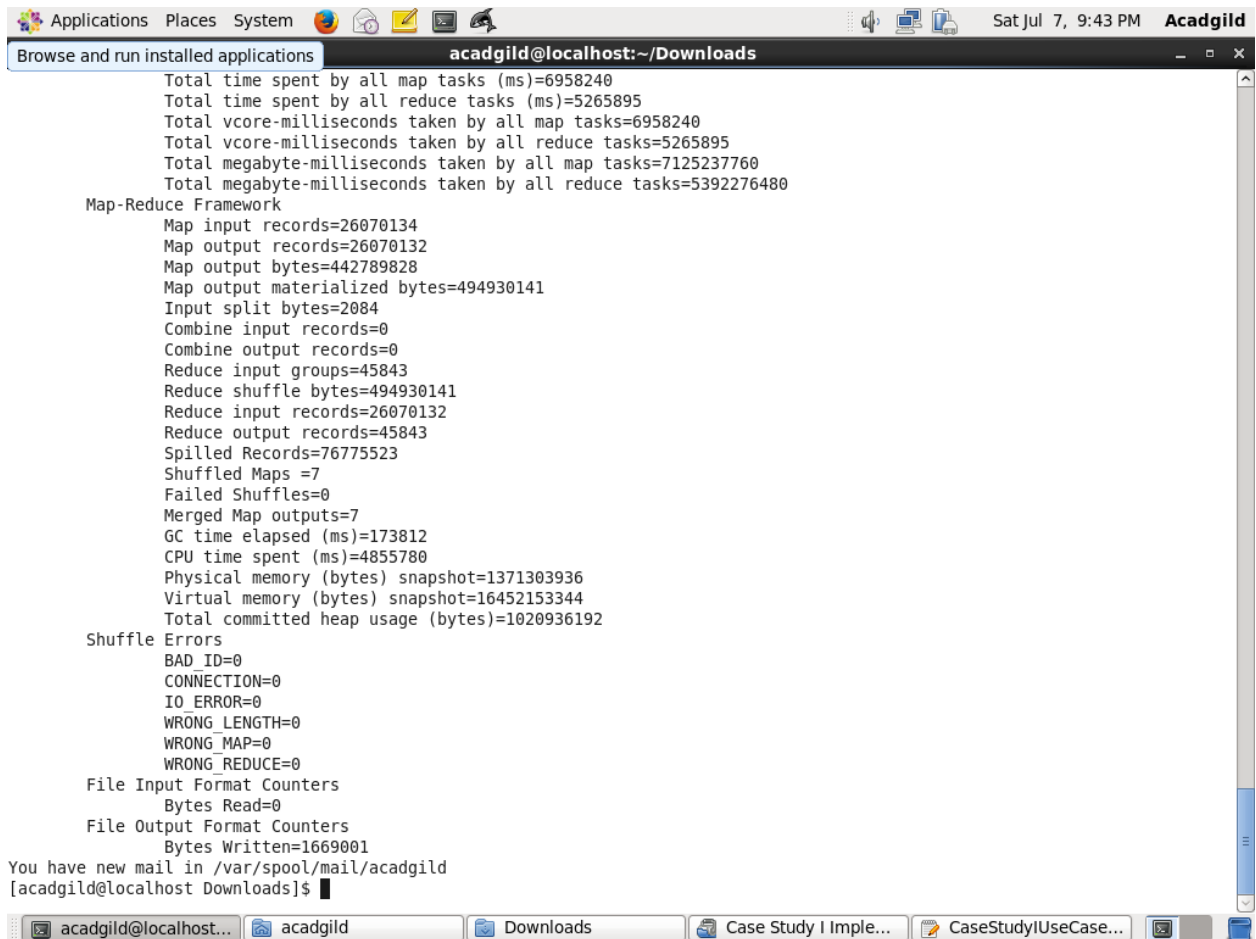
CASESTUDY1 ASSIGNMENT



The screenshot shows a terminal window titled 'acadgild@localhost:~/Downloads'. The window displays the output of the 'hadoop job -list' command, showing counters for a completed job. The counters are organized into three sections: File System Counters, Job Counters, and Map-Reduce Framework. The File System Counters section shows the number of bytes read and written for both the file system and HDFS, as well as the number of read and write operations. The Job Counters section shows the number of map and reduce tasks launched and killed, along with the total time spent by all tasks in various units. The Map-Reduce Framework section shows the number of input and output records, bytes shuffled, and the number of spilled records and failed shuffles.

```
File System Counters
  FILE: Number of bytes read=961694264
  FILE: Number of bytes written=1457487894
  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=1
  Launched map tasks=8
  Launched reduce tasks=1
  Data-local map tasks=8
  Total time spent by all maps in occupied slots (ms)=6958240
  Total time spent by all reduces in occupied slots (ms)=5265895
  Total time spent by all map tasks (ms)=6958240
  Total time spent by all reduce tasks (ms)=5265895
  Total vcore-milliseconds taken by all map tasks=6958240
  Total vcore-milliseconds taken by all reduce tasks=5265895
  Total megabyte-milliseconds taken by all map tasks=7125237760
  Total megabyte-milliseconds taken by all reduce tasks=5392276480
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
```

CASESTUDY1 ASSIGNMENT

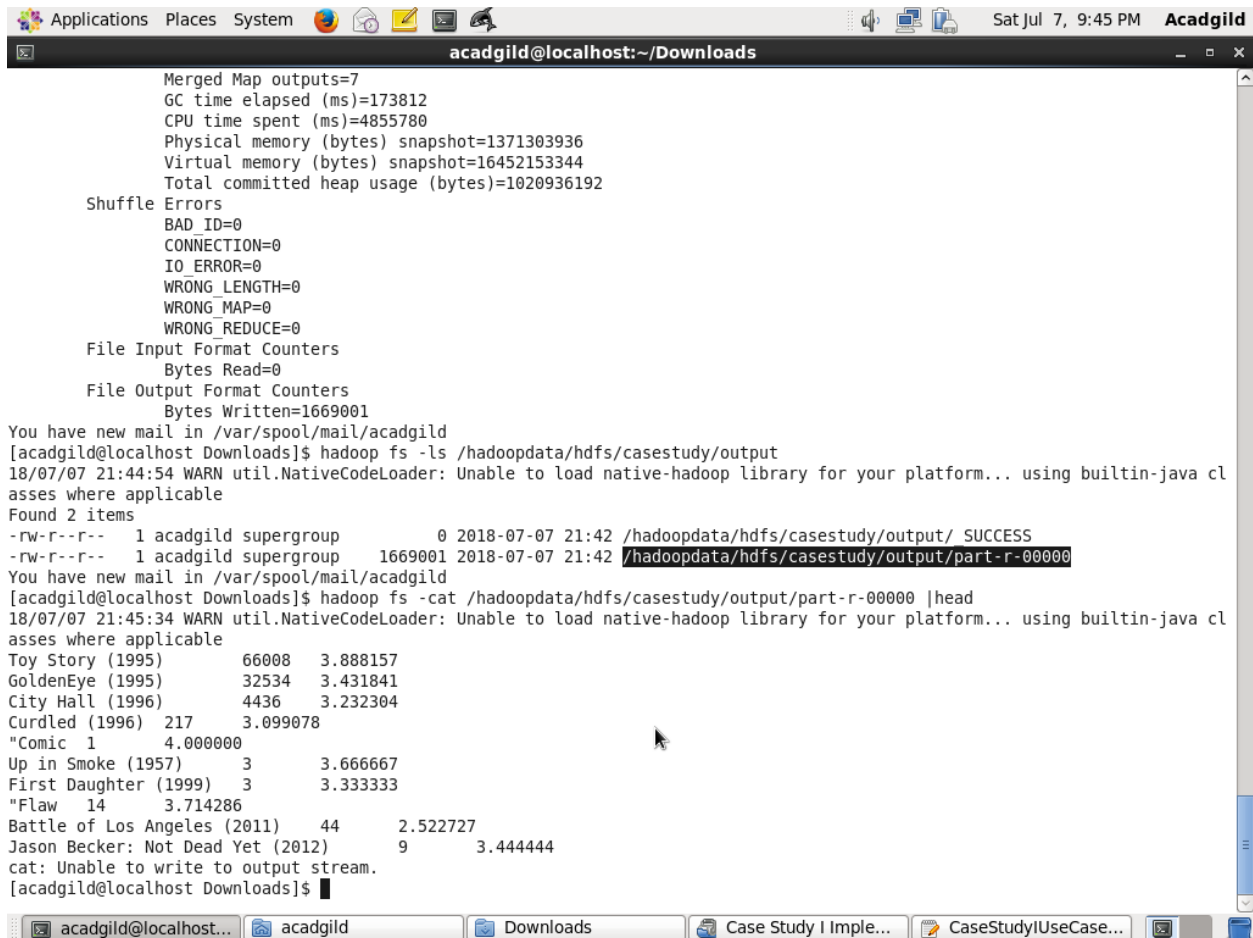


The screenshot shows a terminal window titled 'acadgild@localhost:~/Downloads'. The window displays the output of a Hadoop MapReduce job. The output is organized into sections: overall job statistics, Map-Reduce Framework details, Shuffle Errors, and File Input/Output counters. The job completed successfully with no errors. The terminal window has a standard Ubuntu-style top bar with application, places, and system menus, and a taskbar at the bottom with several open applications.

```
acadgild@localhost:~/Downloads
Total time spent by all map tasks (ms)=6958240
Total time spent by all reduce tasks (ms)=5265895
Total vcore-milliseconds taken by all map tasks=6958240
Total vcore-milliseconds taken by all reduce tasks=5265895
Total megabyte-milliseconds taken by all map tasks=7125237760
Total megabyte-milliseconds taken by all reduce tasks=5392276480
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)=173812
  CPU time spent (ms)=4855780
  Physical memory (bytes) snapshot=1371303936
  Virtual memory (bytes) snapshot=16452153344
  Total committed heap usage (bytes)=1020936192
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
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost Downloads]$
```

CASESTUDY1 ASSIGNMENT

OUTPUT:



```
acadgild@localhost:~/Downloads
Merged Map outputs=7
GC time elapsed (ms)=173812
CPU time spent (ms)=4855780
Physical memory (bytes) snapshot=1371303936
Virtual memory (bytes) snapshot=16452153344
Total committed heap usage (bytes)=1020936192
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
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost Downloads]$ hadoop fs -ls /hadoopdata/hdfs/casestudy/output
18/07/07 21:44:54 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 0 2018-07-07 21:42 /hadoopdata/hdfs/casestudy/output/ SUCCESS
-rw-r--r-- 1 acadgild supergroup 1669001 2018-07-07 21:42 /hadoopdata/hdfs/casestudy/output/part-r-000000
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost Downloads]$ hadoop fs -cat /hadoopdata/hdfs/casestudy/output/part-r-000000 |head
18/07/07 21:45:34 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses 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
Up 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 Downloads]$
```