

Vijay's Assignment – Pig 2

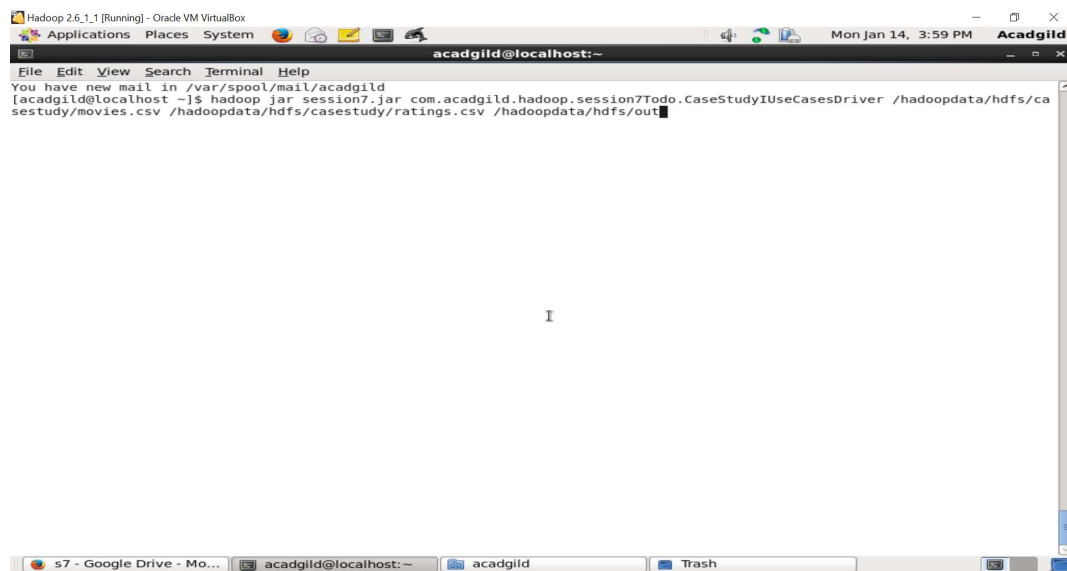
- 1) Session7.jar copied to /home/acadgild folder
- 2) The data files from Google Drive extracted to local drive
- 3) /hadoopdata/hdfs/casestudy path created in HDFS
- 4) All the CSV files received from the googledrive moved to thie HDFS path created above
- 5) Using hadoop jar option "session7.jar" executed to extract the information required below.

Task 1

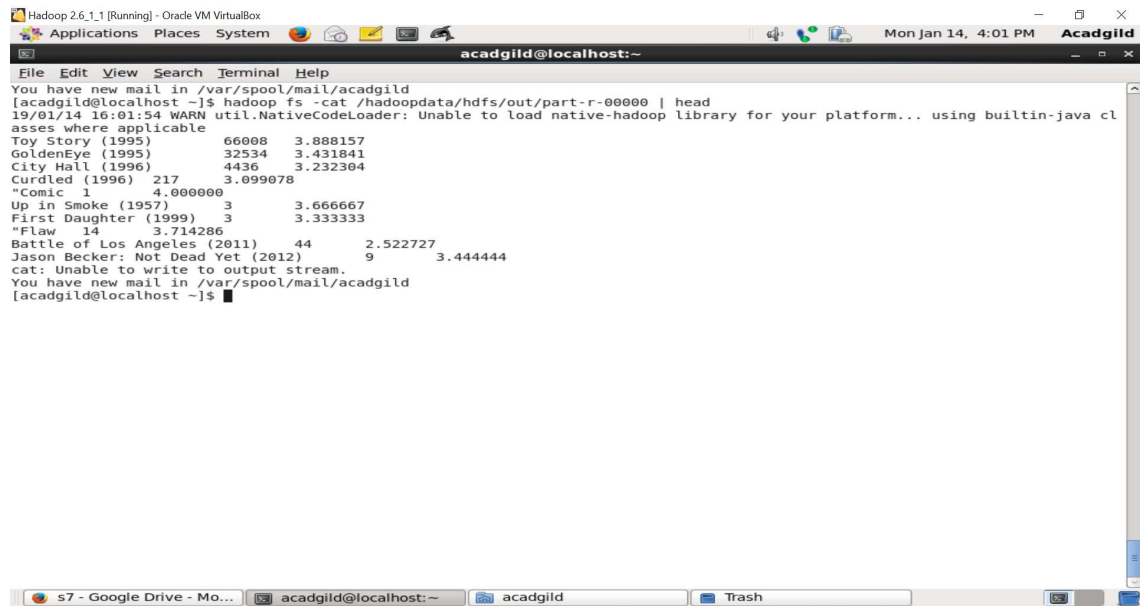
What are the movie titles that the user has rated?

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

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



Output



The screenshot shows a terminal window titled "Hadoop 2.6.1.1 [Running] - Oracle VM VirtualBox". The window has a menu bar with "File", "Edit", "View", "Search", "Terminal", and "Help". The terminal output shows a user prompt, a command to run Hadoop fs, a warning message, and a list of movie titles with their corresponding counts and scores. The terminal also shows a message about new mail and a prompt for the user to enter a command.

```
acacgild@localhost:~$ hadoop fs -cat /hadoopdata/hdfs/out/part-r-00000 | head
19/01/14 16:01:54 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.
You have new mail in /var/spool/mail/acacgild
[acacgild@localhost ~]$
```