

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?

Solution:

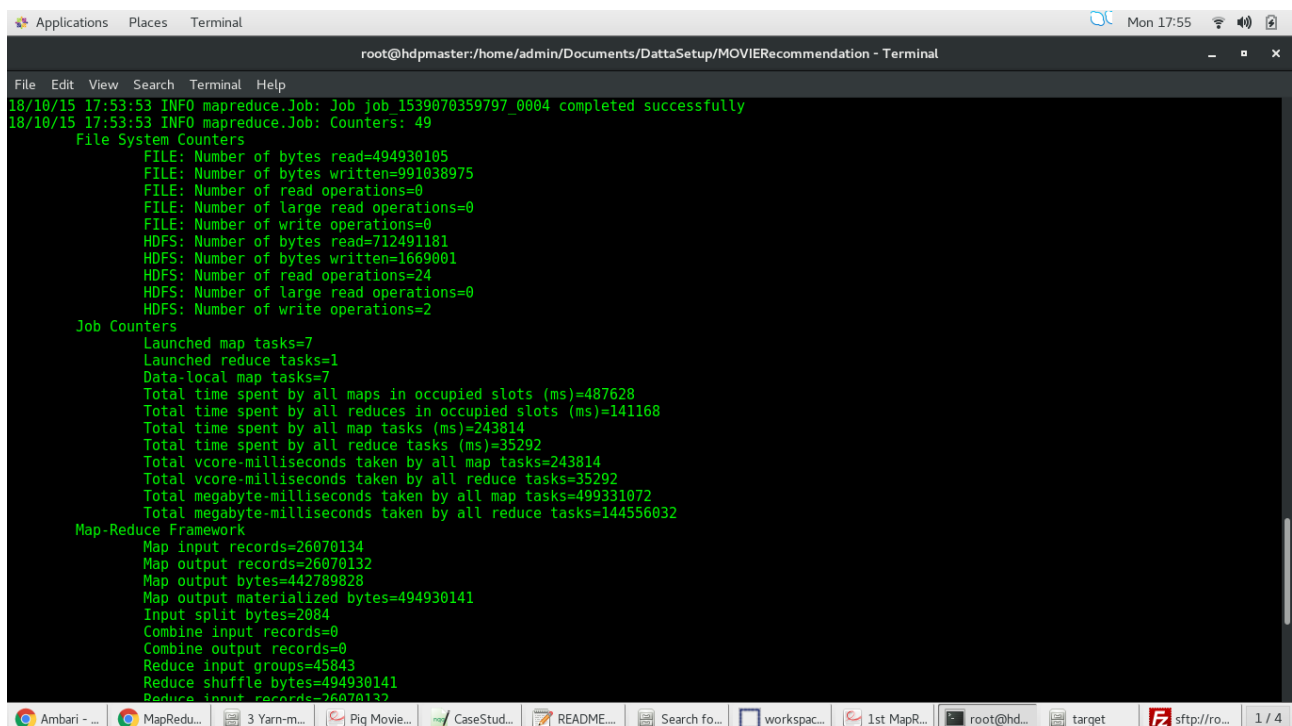
```
[root@hdpmaster MOVIERecommendation]# hadoop jar MovieRatingAssignment-0.0.1-SNAPSHOT.jar Acadgild.MovieRatingAssignment.MovieDriver /user/root/movierecommendation/movies.csv /user/root/movierecommendation/ratings.csv /user/root/movierecommendation/output1/
18/10/15 17:50:42 INFO client.RMProxy: Connecting to ResourceManager at hdpmaster.hortonworks.com/192.168.11.201:8050
18/10/15 17:50:42 INFO client.AHSProxy: Connecting to Application History server at hdpmaster.hortonworks.com/192.168.11.201:10200
18/10/15 17:50:42 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
18/10/15 17:50:43 INFO input.FileInputFormat: Total input paths to process : 1
18/10/15 17:50:43 INFO input.FileInputFormat: Total input paths to process : 1
18/10/15 17:50:43 INFO mapreduce.JobSubmitter: number of splits:7
18/10/15 17:50:43 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1539070359797_0004
18/10/15 17:50:43 INFO impl.YarnClientImpl: Submitted application application_1539070359797_0004
18/10/15 17:50:43 INFO mapreduce.Job: The url to track the job: http://hdpmaster.hortonworks.com:8088/proxy/application_1539070359797_0004/
18/10/15 17:50:43 INFO mapreduce.Job: Running job: job_1539070359797_0004
18/10/15 17:50:52 INFO mapreduce.Job: Job job_1539070359797_0004 running in uber mode : false
18/10/15 17:50:52 INFO mapreduce.Job: map 0% reduce 0%
18/10/15 17:51:50 INFO mapreduce.Job: map 15% reduce 0%
18/10/15 17:51:53 INFO mapreduce.Job: map 19% reduce 0%
18/10/15 17:52:01 INFO mapreduce.Job: map 24% reduce 0%
18/10/15 17:52:02 INFO mapreduce.Job: map 29% reduce 0%
18/10/15 17:52:36 INFO mapreduce.Job: map 42% reduce 0%
18/10/15 17:52:40 INFO mapreduce.Job: map 48% reduce 0%
18/10/15 17:52:46 INFO mapreduce.Job: map 57% reduce 0%
18/10/15 17:52:59 INFO mapreduce.Job: map 78% reduce 0%
18/10/15 17:53:02 INFO mapreduce.Job: map 81% reduce 0%
18/10/15 17:53:08 INFO mapreduce.Job: map 86% reduce 0%
18/10/15 17:53:09 INFO mapreduce.Job: map 100% reduce 0%
18/10/15 17:53:20 INFO mapreduce.Job: map 100% reduce 44%
18/10/15 17:53:24 INFO mapreduce.Job: map 100% reduce 52%
18/10/15 17:53:27 INFO mapreduce.Job: map 100% reduce 60%
18/10/15 17:53:30 INFO mapreduce.Job: map 100% reduce 67%
18/10/15 17:53:33 INFO mapreduce.Job: map 100% reduce 73%
18/10/15 17:53:36 INFO mapreduce.Job: map 100% reduce 79%
18/10/15 17:53:39 INFO mapreduce.Job: map 100% reduce 86%
18/10/15 17:53:43 INFO mapreduce.Job: map 100% reduce 93%
18/10/15 17:53:46 INFO mapreduce.Job: map 100% reduce 100%
```

Run

`hadoop jar MovieRatingAssignment-0.0.1-SNAPSHOT.jar`

`Acadgild.MovieRatingAssignment.MovieDriver /user/root/movierecommendation/movies.csv`

`/user/root/movierecommendation/ratings.csv /user/root/movierecommendation/output1/`



```
18/10/15 17:53:53 INFO mapreduce.Job: Job job_1539070359797_0004 completed successfully
18/10/15 17:53:53 INFO mapreduce.Job: Counters: 49
  File System Counters
    FILE: Number of bytes read=494930105
    FILE: Number of bytes written=991038975
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=712491181
    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
    Launched map tasks=7
    Launched reduce tasks=1
    Data-local map tasks=7
    Total time spent by all maps in occupied slots (ms)=487628
    Total time spent by all reduces in occupied slots (ms)=141168
    Total time spent by all map tasks (ms)=243814
    Total time spent by all reduce tasks (ms)=35292
    Total vcore-milliseconds taken by all map tasks=243814
    Total vcore-milliseconds taken by all reduce tasks=35292
    Total megabyte-milliseconds taken by all map tasks=499331072
    Total megabyte-milliseconds taken by all reduce tasks=144556032
  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
```

```

[root@hdpmaster MOVIERecommendation]# hdfs dfs -ls /user/root/
Found 6 items
drwx----- - root hdfs      0 2018-10-15 17:53 /user/root/.staging
drwxr-xr-x - root hdfs      0 2018-10-15 17:50 /user/root/movierecommendation
-rwxrwxrwx 3 root hdfs    733 2018-10-09 13:17 /user/root/television.txt
drwxr-xr-x - root hdfs      0 2018-10-09 14:26 /user/root/tvdataset1
drwxr-xr-x - root hdfs      0 2018-10-09 16:07 /user/root/tvdataset2
drwxr-xr-x - root hdfs      0 2018-10-09 17:08 /user/root/tvdataset3
[root@hdpmaster MOVIERecommendation]# hdfs dfs -ls /user/root/movierecommendation/
Found 3 items
-rw-r--r-- 3 root hdfs  2283410 2018-10-15 17:49 /user/root/movierecommendation/movies.csv
drwxr-xr-x - root hdfs      0 2018-10-15 17:53 /user/root/movierecommendation/output1
-rw-r--r-- 3 root hdfs  709550327 2018-10-15 17:49 /user/root/movierecommendation/ratings.csv
[root@hdpmaster MOVIERecommendation]# hdfs dfs -ls /user/root/movierecommendation/output1
Found 2 items
-rw-r--r-- 3 root hdfs      0 2018-10-15 17:53 /user/root/movierecommendation/output1/_SUCCESS
-rw-r--r-- 3 root hdfs  1669001 2018-10-15 17:53 /user/root/movierecommendation/output1/part-r-00000
[root@hdpmaster MOVIERecommendation]# █

```

```

[root@hdpmaster MOVIERecommendation]# hdfs dfs -cat /user/root/movierecommendation/output1/part-r-00000 |head -100
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
Chicago Massacre: Richard Speck (2007) 2     2.500000
Keep the Lights On (2012) 25     3.100000
Beauty Is Embarrassing (2012) 15     3.600000
Girl Model (2011)    32     3.281250
Crossfire Hurricane (2012) 18     3.388889
Middle of Nowhere (2012) 11     3.454545
True Blue (2001)     3       3.000000

```

Source of project is attached.