

## Case study One

### 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?

```
Player ▾ | [Icons] | [System Tray]
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$ hadoop jar caseonejar.jar casestudyone.CaseStudyIUseCasesDriver /caseinput/movies.csv /caseinput/ratings.csv /caseoutput/caseout1
18/08/04 08:00:41 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/08/04 08:00:56 INFO client.RMPProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/08/04 08:01:08 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application
ToolRunner to remedy this.
18/08/04 08:01:11 INFO input.FileInputFormat: Total input paths to process : 1
18/08/04 08:01:12 INFO input.FileInputFormat: Total input paths to process : 1
18/08/04 08:01:12 INFO mapreduce.JobSubmitter: number of splits:7
18/08/04 08:01:14 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1533349119711_0001
18/08/04 08:01:17 INFO impl.YarnClientImpl: Submitted application application_1533349119711_0001
18/08/04 08:01:19 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1533349119711_0001/
18/08/04 08:01:19 INFO mapreduce.Job: Running job: job_1533349119711_0001
18/08/04 08:02:55 INFO mapreduce.Job: Job job_1533349119711_0001 running in uber mode : false
18/08/04 08:02:55 INFO mapreduce.Job: map 0% reduce 0%
18/08/04 08:07:34 INFO mapreduce.Job: map 1% reduce 0%
18/08/04 08:07:38 INFO mapreduce.Job: map 2% reduce 0%
18/08/04 08:07:42 INFO mapreduce.Job: map 3% reduce 0%
18/08/04 08:07:48 INFO mapreduce.Job: map 4% reduce 0%
18/08/04 08:07:54 INFO mapreduce.Job: map 5% reduce 0%
18/08/04 08:07:55 INFO mapreduce.Job: map 6% reduce 0%
18/08/04 08:08:00 INFO mapreduce.Job: map 7% reduce 0%
18/08/04 08:08:04 INFO mapreduce.Job: map 8% reduce 0%
18/08/04 08:08:07 INFO mapreduce.Job: map 9% reduce 0%
18/08/04 08:08:09 INFO mapreduce.Job: map 10% reduce 0%
18/08/04 08:08:11 INFO mapreduce.Job: map 11% reduce 0%
18/08/04 08:08:14 INFO mapreduce.Job: map 12% reduce 0%
18/08/04 08:08:15 INFO mapreduce.Job: map 13% reduce 0%
18/08/04 08:08:18 INFO mapreduce.Job: map 16% reduce 0%
18/08/04 08:08:19 INFO mapreduce.Job: map 17% reduce 0%
18/08/04 08:08:21 INFO mapreduce.Job: map 18% reduce 0%
18/08/04 08:08:22 INFO mapreduce.Job: map 21% reduce 0%
18/08/04 08:08:24 INFO mapreduce.Job: map 22% reduce 0%
18/08/04 08:08:25 INFO mapreduce.Job: map 25% reduce 0%
18/08/04 08:08:27 INFO mapreduce.Job: map 26% reduce 0%
18/08/04 08:08:28 INFO mapreduce.Job: map 27% reduce 0%
<
Activate Windows
```

```
Player ▾ | [Icons] | [System Tray]
8/08/04 08:16:05 INFO mapreduce.Job: map 79% reduce 0%
8/08/04 08:16:09 INFO mapreduce.Job: map 80% reduce 0%
8/08/04 08:16:11 INFO mapreduce.Job: map 89% reduce 0%
8/08/04 08:16:13 INFO mapreduce.Job: map 94% reduce 0%
8/08/04 08:16:16 INFO mapreduce.Job: map 95% reduce 0%
8/08/04 08:16:22 INFO mapreduce.Job: map 96% reduce 0%
8/08/04 08:16:33 INFO mapreduce.Job: map 97% reduce 0%
8/08/04 08:16:43 INFO mapreduce.Job: map 98% reduce 0%
8/08/04 08:16:50 INFO mapreduce.Job: map 99% reduce 0%
8/08/04 08:17:00 INFO mapreduce.Job: map 100% reduce 0%
8/08/04 08:18:18 INFO mapreduce.Job: map 100% reduce 14%
8/08/04 08:18:21 INFO mapreduce.Job: map 100% reduce 19%
8/08/04 08:18:24 INFO mapreduce.Job: map 100% reduce 24%
8/08/04 08:18:27 INFO mapreduce.Job: map 100% reduce 29%
8/08/04 08:18:30 INFO mapreduce.Job: map 100% reduce 62%
8/08/04 08:18:34 INFO mapreduce.Job: map 100% reduce 67%
8/08/04 08:18:47 INFO mapreduce.Job: map 100% reduce 68%
8/08/04 08:19:41 INFO mapreduce.Job: map 100% reduce 70%
8/08/04 08:19:53 INFO mapreduce.Job: map 100% reduce 71%
8/08/04 08:20:12 INFO mapreduce.Job: map 100% reduce 72%
8/08/04 08:20:51 INFO mapreduce.Job: map 100% reduce 73%
8/08/04 08:21:10 INFO mapreduce.Job: map 100% reduce 74%
8/08/04 08:21:28 INFO mapreduce.Job: map 100% reduce 75%
8/08/04 08:21:40 INFO mapreduce.Job: map 100% reduce 76%
8/08/04 08:22:00 INFO mapreduce.Job: map 100% reduce 77%
8/08/04 08:22:15 INFO mapreduce.Job: map 100% reduce 78%
8/08/04 08:22:36 INFO mapreduce.Job: map 100% reduce 79%
8/08/04 08:22:59 INFO mapreduce.Job: map 100% reduce 80%
8/08/04 08:23:27 INFO mapreduce.Job: map 100% reduce 81%
8/08/04 08:23:39 INFO mapreduce.Job: map 100% reduce 82%
8/08/04 08:24:01 INFO mapreduce.Job: map 100% reduce 83%
8/08/04 08:24:16 INFO mapreduce.Job: map 100% reduce 84%
8/08/04 08:24:33 INFO mapreduce.Job: map 100% reduce 85%
8/08/04 08:24:50 INFO mapreduce.Job: map 100% reduce 86%
8/08/04 08:25:10 INFO mapreduce.Job: map 100% reduce 87%
8/08/04 08:25:56 INFO mapreduce.Job: map 100% reduce 88%
Activate Windows
```

```
18/08/04 08:28:20 INFO mapreduce.Job: map 100% reduce 97%
18/08/04 08:28:52 INFO mapreduce.Job: map 100% reduce 98%
18/08/04 08:29:39 INFO mapreduce.Job: map 100% reduce 99%
18/08/04 08:30:03 INFO mapreduce.Job: map 100% reduce 100%
18/08/04 08:30:23 INFO mapreduce.Job: Job job_1533349119711_0001 completed successfully
18/08/04 08:30:28 INFO mapreduce.Job: Counters: 50
```

#### File System Counters

```
FILE: Number of bytes read=961694264
FILE: Number of bytes written=1457486766
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=711856049
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)=4980173
Total time spent by all reduces in occupied slots (ms)=830210
Total time spent by all map tasks (ms)=4980173
Total time spent by all reduce tasks (ms)=830210
Total vcore-milliseconds taken by all map tasks=4980173
Total vcore-milliseconds taken by all reduce tasks=830210
Total megabyte-milliseconds taken by all map tasks=5099697152
Total megabyte-milliseconds taken by all reduce tasks=850135040
```

#### Map-Reduce Framework

```
Map input records=26070134
Map output records=26070132
Map output bytes=442789828
Map output materialized bytes=494930141
```

```
Total vcore-milliseconds taken by all reduce tasks=830210
Total megabyte-milliseconds taken by all map tasks=5099697152
Total megabyte-milliseconds taken by all reduce tasks=850135040
```

#### Map-Reduce Framework

```
Map input records=26070134
Map output records=26070132
Map output bytes=442789828
Map output materialized bytes=494930141
Input split bytes=1832
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)=62161
CPU time spent (ms)=846920
Physical memory (bytes) snapshot=1835028480
Virtual memory (bytes) snapshot=16448868352
Total committed heap usage (bytes)=1333788672
```

#### 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/acadnild

---

acadgild@localhost					
File	Edit	View	Search	Terminal	Help
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
"Guns of Fort Petticoat 3				3	3.333333
Human Planet (2011)				197	4.271574
Madagascar (2011)				26	3.769231
Omar Killed Me (Omar m'a tuer) (2011)				9	3.166667
Enola Gay and the Atomic Bombing of Japan (1995)				1	3.500000
Red Hook Summer (2012)				11	2.045455
Stella Maris (1918)				2	3.750000
Die (2010)				10	2.550000
Patrice O'Neal: Elephant in the Room (2011)				22	3.204545
Sunny (Sseo-ni) (2011)				26	3.576923
My Way (Mai Wei) (2011)				30	3.716667
Comme un chef (2012)				83	3.626506
Punching the Clown (2009)				12	3.541667
Metsän tarina (2012)				8	2.875000
Choose (2010)				19	2.815789

```

Texas Across the River (1966) 6 3.500000
[acadgild@localhost ~]$ hadoop fs -cat /caseoutput/caseout1/part-r-00000|head
18/08/04 09:14:26 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
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 ~]$ █

```