Session 7 : Case Study 1
Movie Rating

To-do

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?

Below are Input files path :

1. /hadoopdata/hdfs/caseStudy/movies.csv

2. /hadoopdata/hdfs/caseStudy/ratings.csv

Output file path: **/caseStudy1_Output**

We have created Jar file **caseStudy1_session7.jar** by using java files.

Syntax of command to be used to execute Jar file :

hadoop jar <path of jar file> <Input 1 path> <Input 2 path> <Output Path>

We have used below command to execute Jar file :

hadoop jar caseStudy1_session7.jar   /hadoopdata/hdfs/caseStudy/movies.csv /hadoopdata/hdfs/caseStudy/ratings.csv   /caseStudy1_Output

Session 7 : Case Study 1
Movie Rating

```
[acadgild@localhost ~]$ hadoop jar caseStudy1_session7.jar    /hadoopdata/hdfs/caseStudy/movies.csv /hadoopdata/hdfs/caseStudy/ratings.csv
/caseStudy1_Output
18/08/07 20:01:11 WARN util.NativeCodeLoader: unable to load native-hadoop library for your platform... using builtin-java classes where
applicable
18/08/07 20:01:14 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/08/07 20:01:17 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and
execute your application with ToolRunner to remedy this.
18/08/07 20:01:18 INFO input.FileInputFormat: Total input paths to process : 1
18/08/07 20:01:18 INFO input.FileInputFormat: Total input paths to process : 1
18/08/07 20:01:18 INFO mapreduce.JobSubmitter: number of splits:7
18/08/07 20:01:19 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1533651442083_0001
18/08/07 20:01:20 INFO impl.YarnClientImpl: Submitted application application_1533651442083_0001
18/08/07 20:01:20 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1533651442083_0001/
18/08/07 20:01:20 INFO mapreduce.Job: Running job: job_1533651442083_0001
18/08/07 20:01:43 INFO mapreduce.Job: Job job_1533651442083_0001 running in uber mode : false
18/08/07 20:01:43 INFO mapreduce.Job:   map 0% reduce 0%
18/08/07 20:02:47 INFO mapreduce.Job:   map 1% reduce 0%
18/08/07 20:02:50 INFO mapreduce.Job:   map 2% reduce 0%
18/08/07 20:02:54 INFO mapreduce.Job:   map 4% reduce 0%
18/08/07 20:02:56 INFO mapreduce.Job:   map 5% reduce 0%
18/08/07 20:02:57 INFO mapreduce.Job:   map 7% reduce 0%
18/08/07 20:03:01 INFO mapreduce.Job:   map 10% reduce 0%
18/08/07 20:03:03 INFO mapreduce.Job:   map 11% reduce 0%
18/08/07 20:03:04 INFO mapreduce.Job:   map 13% reduce 0%
18/08/07 20:03:05 INFO mapreduce.Job:   map 14% reduce 0%
18/08/07 20:03:08 INFO mapreduce.Job:   map 17% reduce 0%
18/08/07 20:03:12 INFO mapreduce.Job:   map 19% reduce 0%
18/08/07 20:03:14 INFO mapreduce.Job:   map 20% reduce 0%
18/08/07 20:03:15 INFO mapreduce.Job:   map 22% reduce 0%
18/08/07 20:03:17 INFO mapreduce.Job:   map 23% reduce 0%
18/08/07 20:03:19 INFO mapreduce.Job:   map 24% reduce 0%
18/08/07 20:03:20 INFO mapreduce.Job:   map 25% reduce 0%
18/08/07 20:03:22 INFO mapreduce.Job:   map 26% reduce 0%
18/08/07 20:03:23 INFO mapreduce.Job:   map 27% reduce 0%
```

```
18/08/07 20:03:25 INFO mapreduce.Job:   map 28% reduce 0%
18/08/07 20:03:26 INFO mapreduce.Job:   map 29% reduce 0%
18/08/07 20:03:29 INFO mapreduce.Job:   map 30% reduce 0%
18/08/07 20:03:30 INFO mapreduce.Job:   map 31% reduce 0%
18/08/07 20:03:32 INFO mapreduce.Job:   map 32% reduce 0%
18/08/07 20:03:33 INFO mapreduce.Job:   map 33% reduce 0%
18/08/07 20:03:36 INFO mapreduce.Job:   map 34% reduce 0%
18/08/07 20:03:39 INFO mapreduce.Job:   map 35% reduce 0%
18/08/07 20:03:42 INFO mapreduce.Job:   map 36% reduce 0%
18/08/07 20:03:46 INFO mapreduce.Job:   map 37% reduce 0%
18/08/07 20:03:49 INFO mapreduce.Job:   map 38% reduce 0%
18/08/07 20:03:50 INFO mapreduce.Job:   map 39% reduce 0%
18/08/07 20:03:54 INFO mapreduce.Job:   map 40% reduce 0%
18/08/07 20:04:01 INFO mapreduce.Job:   map 45% reduce 0%
18/08/07 20:04:48 INFO mapreduce.Job:   map 46% reduce 0%
18/08/07 20:04:51 INFO mapreduce.Job:   map 47% reduce 0%
18/08/07 20:04:52 INFO mapreduce.Job:   map 48% reduce 0%
18/08/07 20:04:54 INFO mapreduce.Job:   map 49% reduce 0%
18/08/07 20:04:56 INFO mapreduce.Job:   map 50% reduce 0%
18/08/07 20:04:59 INFO mapreduce.Job:   map 51% reduce 0%
18/08/07 20:05:00 INFO mapreduce.Job:   map 52% reduce 0%
18/08/07 20:05:05 INFO mapreduce.Job:   map 54% reduce 0%
18/08/07 20:05:09 INFO mapreduce.Job:   map 56% reduce 0%
18/08/07 20:05:12 INFO mapreduce.Job:   map 58% reduce 0%
18/08/07 20:05:13 INFO mapreduce.Job:   map 59% reduce 0%
18/08/07 20:05:15 INFO mapreduce.Job:   map 60% reduce 0%
18/08/07 20:05:16 INFO mapreduce.Job:   map 61% reduce 0%
18/08/07 20:05:18 INFO mapreduce.Job:   map 62% reduce 0%
18/08/07 20:05:30 INFO mapreduce.Job:   map 71% reduce 0%
18/08/07 20:05:32 INFO mapreduce.Job:   map 76% reduce 0%
18/08/07 20:06:06 INFO mapreduce.Job:   map 77% reduce 0%
18/08/07 20:06:13 INFO mapreduce.Job:   map 78% reduce 0%
18/08/07 20:06:16 INFO mapreduce.Job:   map 79% reduce 0%
18/08/07 20:06:19 INFO mapreduce.Job:   map 81% reduce 0%
18/08/07 20:06:23 INFO mapreduce.Job:   map 83% reduce 0%
18/08/07 20:06:26 INFO mapreduce.Job:   map 84% reduce 0%
```

```
18/08/07 20:06:28 INFO mapreduce.Job:  map 85% reduce 5%
18/08/07 20:06:29 INFO mapreduce.Job:  map 86% reduce 5%
18/08/07 20:06:32 INFO mapreduce.Job:  map 88% reduce 5%
18/08/07 20:06:36 INFO mapreduce.Job:  map 90% reduce 5%
18/08/07 20:06:39 INFO mapreduce.Job:  map 92% reduce 5%
18/08/07 20:06:42 INFO mapreduce.Job:  map 94% reduce 5%
18/08/07 20:06:45 INFO mapreduce.Job:  map 96% reduce 5%
18/08/07 20:06:48 INFO mapreduce.Job:  map 98% reduce 5%
18/08/07 20:06:51 INFO mapreduce.Job:  map 99% reduce 5%
18/08/07 20:06:52 INFO mapreduce.Job:  map 100% reduce 7%
18/08/07 20:06:58 INFO mapreduce.Job:  map 100% reduce 20%
18/08/07 20:07:01 INFO mapreduce.Job:  map 100% reduce 33%
18/08/07 20:07:10 INFO mapreduce.Job:  map 100% reduce 34%
18/08/07 20:07:19 INFO mapreduce.Job:  map 100% reduce 64%
18/08/07 20:07:22 INFO mapreduce.Job:  map 100% reduce 67%
18/08/07 20:07:28 INFO mapreduce.Job:  map 100% reduce 68%
18/08/07 20:07:41 INFO mapreduce.Job:  map 100% reduce 69%
18/08/07 20:07:56 INFO mapreduce.Job:  map 100% reduce 70%
18/08/07 20:08:09 INFO mapreduce.Job:  map 100% reduce 71%
18/08/07 20:08:22 INFO mapreduce.Job:  map 100% reduce 72%
18/08/07 20:08:37 INFO mapreduce.Job:  map 100% reduce 73%
18/08/07 20:08:49 INFO mapreduce.Job:  map 100% reduce 74%
18/08/07 20:09:02 INFO mapreduce.Job:  map 100% reduce 75%
18/08/07 20:09:16 INFO mapreduce.Job:  map 100% reduce 76%
18/08/07 20:09:28 INFO mapreduce.Job:  map 100% reduce 77%
18/08/07 20:09:41 INFO mapreduce.Job:  map 100% reduce 78%
18/08/07 20:09:53 INFO mapreduce.Job:  map 100% reduce 79%
18/08/07 20:10:06 INFO mapreduce.Job:  map 100% reduce 80%
18/08/07 20:10:20 INFO mapreduce.Job:  map 100% reduce 81%
18/08/07 20:10:32 INFO mapreduce.Job:  map 100% reduce 82%
18/08/07 20:10:45 INFO mapreduce.Job:  map 100% reduce 83%
18/08/07 20:10:57 INFO mapreduce.Job:  map 100% reduce 84%
18/08/07 20:11:10 INFO mapreduce.Job:  map 100% reduce 85%
18/08/07 20:11:22 INFO mapreduce.Job:  map 100% reduce 86%
18/08/07 20:11:37 INFO mapreduce.Job:  map 100% reduce 87%
18/08/07 20:11:49 INFO mapreduce.Job:  map 100% reduce 88%
18/08/07 20:12:02 INFO mapreduce.Job:  map 100% reduce 89%
```

```
18/08/07 20:12:17 INFO mapreduce.Job:  map 100% reduce 90%
18/08/07 20:12:30 INFO mapreduce.Job:  map 100% reduce 91%
18/08/07 20:12:44 INFO mapreduce.Job:  map 100% reduce 92%
18/08/07 20:12:57 INFO mapreduce.Job:  map 100% reduce 93%
18/08/07 20:13:09 INFO mapreduce.Job:  map 100% reduce 94%
18/08/07 20:13:22 INFO mapreduce.Job:  map 100% reduce 95%
18/08/07 20:13:36 INFO mapreduce.Job:  map 100% reduce 96%
18/08/07 20:13:48 INFO mapreduce.Job:  map 100% reduce 97%
18/08/07 20:14:01 INFO mapreduce.Job:  map 100% reduce 98%
18/08/07 20:14:13 INFO mapreduce.Job:  map 100% reduce 99%
18/08/07 20:14:28 INFO mapreduce.Job:  map 100% reduce 100%
18/08/07 20:14:35 INFO mapreduce.Job: Job job_1533651442083_0001 completed successfully
18/08/07 20:14:37 INFO mapreduce.Job: Counters: 50
        File System Counters
                FILE: Number of bytes read=961719024
                FILE: Number of bytes written=1457594999
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=711856070
                HDFS: Number of bytes written=1669001
                HDFS: Number of read operations=27
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=4
        Job Counters
                Killed map tasks=1
                Launched map tasks=7
                Launched reduce tasks=2
                Data-local map tasks=7
                Total time spent by all maps in occupied slots (ms)=1750715
                Total time spent by all reduces in occupied slots (ms)=1001843
                Total time spent by all map tasks (ms)=1750715
                Total time spent by all reduce tasks (ms)=1001843
                Total vcore-milliseconds taken by all map tasks=1750715
                Total vcore-milliseconds taken by all reduce tasks=1001843
                Total megabyte-milliseconds taken by all map tasks=1792732160
```

```
                Total megabyte-milliseconds taken by all map tasks=1792732160
                Total megabyte-milliseconds taken by all reduce tasks=1025887232
        Map-Reduce Framework
                Map input records=26070134
                Map output records=26070132
                Map output bytes=442789828
                Map output materialized bytes=494930183
                Input split bytes=1853
                Combine input records=0
                Combine output records=0
                Reduce input groups=45843
                Reduce shuffle bytes=494930183
                Reduce input records=26070132
                Reduce output records=45843
                Spilled Records=76775523
                Shuffled Maps =14
                Failed Shuffles=0
                Merged Map outputs=14
                GC time elapsed (ms)=20555
                CPU time spent (ms)=612350
                Physical memory (bytes) snapshot=1559117824
                Virtual memory (bytes) snapshot=18513768448
                Total committed heap usage (bytes)=1047478272
        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 ~]$ hadoop fs -ls /caseStudy1_Output
18/08/07 20:15:33 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where
applicable
Found 3 items
-rw-r--r--   1 acadgild supergroup          0 2018-08-07 20:14 /caseStudy1_Output/_SUCCESS
-rw-r--r--   1 acadgild supergroup     839548 2018-08-07 20:14 /caseStudy1_Output/part-r-00000
```

Please refer below output :

First field shows list of movie titles that user has rated.

Second field shows number of times a movie has been rated by user.

Third field shows Average Rating given for movie.

e.g. Movie '**Toy Story (1995)**' has been rated by users by **66008** times and Average Rating given for this movie is **3.888157**.

```
[acadgild@localhost ~]$ hadoop fs -cat /caseStudy1_Output/part-r-00000 | head
18/08/07 20:19:29 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
First Daughter (1999)   3       3.333333
"Flaw   14      3.714286
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
True Blue (2001)        3       3.000000
cat: Unable to write to output stream.
You have new mail in /var/spool/mail/acadgild
```