

## Case Study 1

### 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?

Note: Used a different dataset given by mentor which is of less size than the dataset given in the case study. Used different dataset as the mapreduce program is taking too much time and getting struck midway.

Used the files highlighted in red below:

```
[acadgild@localhost ~]$ hadoop fs -ls /hadoopdata
18/05/17 22:37:58 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 8 items
-rw-r--r-- 1 acadgild supergroup      12 2018-05-04 16:44 /hadoopdata/A.txt
-rw-r--r-- 1 acadgild supergroup 2283410 2018-05-17 21:50 /hadoopdata/movies.csv
-rw-r--r-- 1 acadgild supergroup 458371 2018-05-17 22:37 /hadoopdata/movies.txt
-rw-r--r-- 1 acadgild supergroup    72 2018-05-04 16:59 /hadoopdata/musicdata.txt
-rw-r--r-- 1 acadgild supergroup 709550327 2018-05-17 21:49 /hadoopdata/ratings.csv
-rw-r--r-- 1 acadgild supergroup 2438233 2018-05-17 22:36 /hadoopdata/ratings.txt
-rw-r--r-- 1 acadgild supergroup    733 2018-05-07 11:26 /hadoopdata/television.txt
-rw-r--r-- 1 acadgild supergroup    82 2018-05-04 16:42 /hadoopdata/test.txt
```

The source code mentioned below is uploaded separately with code explanation

Driver class file - ReduceJoin.java

Movie Mapper class file - MovieMapper.java

Rating Mapper class file - RatingMapper.java

Reducer class file - ReduceJoinReducer.java

Command to execute Mapreduce:

**`hadoop jar MovieRating.jar /hadoopdata/movies.txt`** Command to run the mapreduce:  
**`/hadoopdata/ratings.txt /MovieRatingOutput`**

```

[acadgild@localhost ~]$ hadoop jar MovieRating.jar /hadoopdata/movies.txt /hadoopdata/ratings.txt /MovieRatingOutput
Config setting is done
18/05/17 22:49:27 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
job setting is done
setJarByClass setting is done
Added input paths setting is done
Reducer class setting is done
output key values is done
outputpath setting is done
18/05/17 22:49:29 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/05/17 22:49:30 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
18/05/17 22:49:31 INFO input.FileInputFormat: Total input paths to process : 1
18/05/17 22:49:31 INFO input.FileInputFormat: Total input paths to process : 1
18/05/17 22:49:31 INFO mapreduce.JobSubmitter: number of splits:2
18/05/17 22:49:31 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1526577508117_0001
18/05/17 22:49:32 INFO impl.YarnClientImpl: Submitted application application_1526577508117_0001
18/05/17 22:49:32 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1526577508117_0001/
18/05/17 22:49:32 INFO mapreduce.Job: Running job: job_1526577508117_0001
18/05/17 22:49:44 INFO mapreduce.Job: Job job_1526577508117_0001 running in uber mode : false
18/05/17 22:49:44 INFO mapreduce.Job: map 0% reduce 0%
18/05/17 22:49:52 INFO mapreduce.Job: map 50% reduce 0%
18/05/17 22:49:54 INFO mapreduce.Job: map 100% reduce 0%
18/05/17 22:50:01 INFO mapreduce.Job: map 100% reduce 100%
18/05/17 22:50:02 INFO mapreduce.Job: Job job_1526577508117_0001 completed successfully

```

## Output after running the mapreduce:

```

[acadgild@localhost ~]$ hadoop fs -ls /MovieRatingOutput
18/05/17 22:53:37 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 2 items
-rw-r--r-- 1 acadgild supergroup 0 2018-05-17 22:50 /MovieRatingOutput/_SUCCESS
-rw-r--r-- 1 acadgild supergroup 315116 2018-05-17 22:50 /MovieRatingOutput/part-r-000000
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$ hadoop fs -cat /MovieRatingOutput/part-r-000000 | head
18/05/17 22:54:54 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Toy Story (1995) 247 3.872470
GoldenEye (1995) 122 3.450820
City Hall (1996) 21 3.428571
Keep the Lights On (2012) 1 3.000000
Beauty Is Embarrassing (2012) 1 2.000000
Girl Model (2011) 1 4.000000
Movie 43 (2013) 2 1.500000
"Pervert's Guide to Ideology 1 4.500000
Sightseers (2012) 2 3.500000
Hansel & Gretel: Witch Hunters (2013) 2 2.250000
cat: Unable to write to output stream.
You have new mail in /var/spool/mail/acadgild

```