CASE_STUDY _I

Movie Rating

PROBLEM STATEMENT -

1. What are the movie titles that the user has rated?
2. How many times a movie has been rated by the user?
3. In question 2 above, what is the average rating given for a movie?

SOLUTION -

Mapper method –

```java
public class CaseStudyIUseCasesMoviesMapper extends
        Mapper<LongWritable, Text, Text, Text> {

public void map(LongWritable key, Text value, Context context)
            throws IOException, InterruptedException {

    try {
        if (key.get() == 0 && value.toString().contains("movieId")){
            return;
        } else {
            String record = value.toString();
            String[] parts = record.split(","); //Split input with delimiter
            context.write(new Text(parts[0]), new Text("movies\t" + parts[1])); //Output of Mapper
        }
```

```java
public void map(LongWritable key, Text value, Context context)
            throws IOException, InterruptedException {

    try {
        if (key.get() == 0 && value.toString().contains("userId")){
            return;
        } else {
            String record = value.toString();
            String[] parts = record.split(",");
            context.write(new Text(parts[1]), new Text("ratings\t" + parts[2]));//Output of Mapper
        }
    } catch (Exception e) {
        e.printStackTrace();
    }
}
```

Reducer method –

```java
public void reduce(Text key, Iterable<Text> values, Context context)
        throws IOException, InterruptedException {
    String titles = "";
    double total = 0.0;
    int count = 0;
    System.out.println("Text Key    =>"+key.toString());
    for (Text t : values) {
        String parts[] = t.toString().split("\t");
        System.out.println("Text values =>"+t.toString());
        if (parts[0].equals("ratings")) {
            count++; // Number of user rating for each movie
            String rating = parts[1].trim();
            System.out.println("Rating is =>"+rating);
            total += Double.parseDouble(rating);
        } else if (parts[0].equals("movies")) {
            titles = parts[1]; //Movie name
        }
    }

    double average = total / count; //Average rating of the movie
    String str = String.format("%d\t%f", count, average);
    context.write(new Text(titles), new Text(str));
}
```

OUTPUT –

Command to run the jar –

**hadoop jar MapReduceCaseStudy.jar
CaseStudy.CaseStudyIUseCasesDriver /CaseStudy/movies.csv
/CaseStudy/ratings.csv /MovieRating**

Input files - **movies.csv   ratings.csv**

Output directoy - **MovieRating**

```
[acadgild@localhost hadoop]$ hadoop jar MapReduceCaseStudy.jar CaseStudy.CaseStudyIUseCasesDriver /CaseStudy/movies.csv /CaseStudy/ratings.csv /MovieRating
18/09/26 01:31:11 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/09/26 01:31:15 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/09/26 01:31:16 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application wi
th ToolRunner to remedy this.
18/09/26 01:31:17 INFO input.FileInputFormat: Total input paths to process : 1
18/09/26 01:31:17 INFO input.FileInputFormat: Total input paths to process : 1
18/09/26 01:31:18 INFO mapreduce.JobSubmitter: number of splits:7
18/09/26 01:31:18 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1537895912006_0003
18/09/26 01:31:19 INFO impl.YarnClientImpl: Submitted application application_1537895912006_0003
18/09/26 01:31:19 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1537895912006_0003/
```

```
[acadgild@localhost hadoop]$ hadoop jar MapReduceCaseStudy.jar CaseStudy.CaseStudyIUseCasesDriver /CaseStudy/movies.csv /CaseStudy/ratings.csv /MovieRating
18/09/26 01:31:11 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/09/26 01:31:15 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/09/26 01:31:16 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application wi
th ToolRunner to remedy this.
18/09/26 01:31:17 INFO input.FileInputFormat: Total input paths to process : 1
18/09/26 01:31:17 INFO input.FileInputFormat: Total input paths to process : 1
18/09/26 01:31:18 INFO mapreduce.JobSubmitter: number of splits:7
18/09/26 01:31:18 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1537895912006_0003
18/09/26 01:31:19 INFO impl.YarnClientImpl: Submitted application application_1537895912006_0003
18/09/26 01:31:19 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1537895912006_0003/
18/09/26 01:31:19 INFO mapreduce.Job: Running job: job_1537895912006_0003
18/09/26 01:31:39 INFO mapreduce.Job: Job job_1537895912006_0003 running in uber mode : false
18/09/26 01:31:39 INFO mapreduce.Job:   map 0% reduce 0%
18/09/26 01:34:36 INFO mapreduce.Job:   map 1% reduce 0%
18/09/26 01:34:37 INFO mapreduce.Job:   map 2% reduce 0%
18/09/26 01:34:38 INFO mapreduce.Job:   map 3% reduce 0%
18/09/26 01:34:40 INFO mapreduce.Job:   map 5% reduce 0%
18/09/26 01:34:46 INFO mapreduce.Job:   map 6% reduce 0%
18/09/26 01:34:47 INFO mapreduce.Job:   map 7% reduce 0%
18/09/26 01:35:00 INFO mapreduce.Job:   map 8% reduce 0%
18/09/26 01:35:03 INFO mapreduce.Job:   map 10% reduce 0%
18/09/26 01:35:06 INFO mapreduce.Job:   map 11% reduce 0%
18/09/26 01:35:09 INFO mapreduce.Job:   map 12% reduce 0%
18/09/26 01:35:16 INFO mapreduce.Job:   map 13% reduce 0%
18/09/26 01:35:20 INFO mapreduce.Job:   map 14% reduce 0%
18/09/26 01:35:36 INFO mapreduce.Job:   map 15% reduce 0%
18/09/26 01:35:39 INFO mapreduce.Job:   map 17% reduce 0%
18/09/26 01:35:42 INFO mapreduce.Job:   map 18% reduce 0%
18/09/26 01:35:49 INFO mapreduce.Job:   map 19% reduce 0%
18/09/26 01:36:25 INFO mapreduce.Job:   map 20% reduce 0%
18/09/26 01:36:28 INFO mapreduce.Job:   map 21% reduce 0%
18/09/26 01:36:49 INFO mapreduce.Job:   map 24% reduce 0%
18/09/26 01:36:52 INFO mapreduce.Job:   map 25% reduce 0%
18/09/26 01:37:35 INFO mapreduce.Job:   map 27% reduce 0%
18/09/26 01:37:38 INFO mapreduce.Job:   map 28% reduce 0%
```

```
18/09/26 02:24:26 INFO mapreduce.Job:   map 100% reduce 99%
18/09/26 02:25:40 INFO mapreduce.Job:   map 100% reduce 100%
18/09/26 02:26:21 INFO mapreduce.Job: Job job_1537895912006_0003 completed successfully
18/09/26 02:26:24 INFO mapreduce.Job: Counters: 50
        File System Counters
                FILE: Number of bytes read=961694264
                FILE: Number of bytes written=1457486630
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=711856028
                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)=4500973
                Total time spent by all reduces in occupied slots (ms)=2508422
                Total time spent by all map tasks (ms)=4500973
                Total time spent by all reduce tasks (ms)=2508422
                Total vcore-milliseconds taken by all map tasks=4500973
                Total vcore-milliseconds taken by all reduce tasks=2508422
                Total megabyte-milliseconds taken by all map tasks=4608996352
                Total megabyte-milliseconds taken by all reduce tasks=2568624128
        Map-Reduce Framework
                Map input records=26070134
                Map output records=26070132
                Map output bytes=442789828
                Map output materialized bytes=494930141
                Input split bytes=1811
                Combine input records=0
                Combine output records=0
                Reduce input groups=45843
                Reduce shuffle bytes=494930141
```

Output File –

Output file **part-r-00000** is in the directory **/MovieRating**

```
[acadgild@localhost hadoop]$ hadoop fs -ls /MovieRating
18/09/26 02:33:04 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform...
Found 2 items
-rw-r--r--   1 acadgild supergroup          0 2018-09-26 02:26 /MovieRating/_SUCCESS
-rw-r--r--   1 acadgild supergroup    1669001 2018-09-26 02:26 /MovieRating/part-r-00000
[acadgild@localhost hadoop]$
```

Output –

```
[acadgild@localhost hadoop]$ hadoop fs -cat /MovieRating/part-r-00000
18/09/26 02:35:07 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform...
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       3.333333
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
Made in Hong Kong (Xiang Gang zhi zao) (1997)   3       2.333333
Creature from Black Lake (1976)         1       1.500000
Bad Luck (Zezowate szczescie) (1960)    5       3.700000
```