

### 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?

### Code:

**CaseStudyIUseCasesReducer.java**

**Reducer class is used to arrive at the output. Input is received from Mapper classes of Movies and Rating.**

**Rating Mapper Class returns Movie Id and Rating**

**Movie Mapper Class returns Movie Id and Name**

**Driver Class has all the configurations required .**

```
import java.io.IOException;

import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Reducer;

public class CaseStudyIUseCasesReducer extends
    Reducer<Text, Text, Text, Text> {

    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++;
                String rating = parts[1].trim();
                System.out.println("Rating is
=>" + rating);
                total +=
                Double.parseDouble(rating);
            } else if (parts[0].equals("movies")) {
                titles = parts[1];
            }
        }

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

```
    }
}
```

### Ratings Mapper Class:

```
import java.io.IOException;

import org.apache.hadoop.io.LongWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Mapper;

public class CaseStudyIUseCasesRatingsMapper 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("userId")) {
                return;
            } else {
                String record = value.toString();
                String[] parts = record.split(",");
                context.write(new Text(parts[1]), new
Text("ratings\t" + parts[2]));
            }
        } catch (Exception e) {
            e.printStackTrace();
        }

    }
}
```

### Movies Mapper Class:

```
import java.io.IOException;

import org.apache.hadoop.io.LongWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Mapper;

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(",");
        context.write(new Text(parts[0]), new
Text("movies\t" + parts[1]));
    }
    } catch (Exception e) {
        e.printStackTrace();
    }
}
}

```

#### Driver Class:

```

import org.apache.hadoop.conf.Configuration;
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.lib.output.TextOutputFormat;
import org.apache.hadoop.mapreduce.lib.input.TextInputFormat;
import org.apache.hadoop.mapreduce.Job;
import org.apache.hadoop.mapreduce.lib.input.MultipleInputs;
import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;

public class CaseStudyIUseCasesDriver {

    @SuppressWarnings("deprecation")
    public static void main(String[] args) throws Exception {
        if (args.length != 3) {
            System.err.println("Usage: CaseStudyIUseCase2Driver <input path1>
<input path2> <output path>");
            System.exit(-1);
        }

        //Job Related Configurations
        Configuration conf = new Configuration();
        Job job = new Job(conf, "CaseStudyIUseCase2Driver");
        job.setJarByClass(CaseStudyIUseCasesDriver.class);

        //Since there are multiple input, there is a slightly different
way of specifying input path, input format and mapper
        MultipleInputs.addInputPath(job, new
Path(args[0]), TextInputFormat.class,
CaseStudyIUseCasesMoviesMapper.class);
        MultipleInputs.addInputPath(job, new
Path(args[1]), TextInputFormat.class,
CaseStudyIUseCasesRatingsMapper.class);

        //Set the reducer
        job.setReducerClass(CaseStudyIUseCasesReducer.class);

        //set the out path
        Path outputPath = new Path(args[2]);
        FileOutputFormat.setOutputPath(job, outputPath);
        outputPath.getFileSystem(conf).delete(outputPath, true);

        //set up the output key and value classes
        job.setOutputKeyClass(Text.class);
        job.setOutputValueClass(Text.class);
    }
}

```

```

    //execute the job
    System.exit(job.waitForCompletion(true) ? 0 : 1);
}
}

```

## OUTPUT

```

[acadgild@localhost ~]$ hadoop jar CaseStudyIUseCasesDriver.jar CaseStudyIUseCasesDriver /user/acadgild/movies.csv /user/acadgild/ratings.csv /user/acadgild/MOVIESOUT
18/08/07 16:25:34 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/08/07 16:25:35 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/08/07 16:25:37 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 16:25:37 INFO input.FileInputFormat: Total input paths to process : 1
18/08/07 16:25:37 INFO input.FileInputFormat: Total input paths to process : 1
18/08/07 16:25:37 INFO mapreduce.JobSubmitter: number of splits:7
18/08/07 16:25:37 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1533637955065_0003
18/08/07 16:25:38 INFO impl.YarnClientImpl: Submitted application application_1533637955065_0003
18/08/07 16:25:38 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1533637955065_0003/
18/08/07 16:25:38 INFO mapreduce.Job: Running job: job_1533637955065_0003
18/08/07 16:25:50 INFO mapreduce.Job: Job job_1533637955065_0003 running in uber mode : false
18/08/07 16:25:50 INFO mapreduce.Job: map 0% reduce 0%
18/08/07 16:26:34 INFO mapreduce.Job: map 1% reduce 0%
18/08/07 16:26:37 INFO mapreduce.Job: map 2% reduce 0%
18/08/07 16:26:39 INFO mapreduce.Job: map 3% reduce 0%
18/08/07 16:26:41 INFO mapreduce.Job: map 5% reduce 0%
18/08/07 16:26:42 INFO mapreduce.Job: map 6% reduce 0%
18/08/07 16:26:44 INFO mapreduce.Job: map 9% reduce 0%
18/08/07 16:26:45 INFO mapreduce.Job: map 10% reduce 0%
18/08/07 16:26:48 INFO mapreduce.Job: map 16% reduce 0%
18/08/07 16:26:51 INFO mapreduce.Job: map 20% reduce 0%
18/08/07 16:26:54 INFO mapreduce.Job: map 25% reduce 0%
18/08/07 16:26:57 INFO mapreduce.Job: map 28% reduce 0%
18/08/07 16:27:00 INFO mapreduce.Job: map 32% reduce 0%
18/08/07 16:27:03 INFO mapreduce.Job: map 33% reduce 0%
18/08/07 16:27:04 INFO mapreduce.Job: map 35% reduce 0%
18/08/07 16:27:08 INFO mapreduce.Job: map 37% reduce 0%
18/08/07 16:27:11 INFO mapreduce.Job: map 38% reduce 0%
18/08/07 16:27:14 INFO mapreduce.Job: map 40% reduce 0%
18/08/07 16:27:17 INFO mapreduce.Job: map 45% reduce 0%
18/08/07 16:27:46 INFO mapreduce.Job: map 59% reduce 0%
18/08/07 16:27:48 INFO mapreduce.Job: map 60% reduce 0%
18/08/07 16:27:49 INFO mapreduce.Job: map 63% reduce 0%
18/08/07 16:27:51 INFO mapreduce.Job: map 69% reduce 0%

[acadgild@localhost ~]$ hdfs dfs -ls /user/acadgild/MOVIESOUT
18/08/07 18:00:24 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 2018-08-07 17:56 /user/acadgild/MOVIESOUT/_SUCCESS
-rw-r--r-- 1 acadgild supergroup 1669001 2018-08-07 17:56 /user/acadgild/MOVIESOUT/part-r-000000
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$

You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$ hdfs dfs -cat /user/acadgild/MOVIESOUT/part-r-000000

```

Applications Places System Tue Aug 7, 6:01 PM Acadgild

acadgild@localhost:~

File Edit View Search Terminal Help

Vamps (2012)	44	2.795455		
Cash on Demand (1962)	2	3.750000		
"Naughty Room	1	3.000000		
"Moth Diaries	39	3.012821		
Joe + Belle (2011)	2	4.000000		
"Invisible War	86	3.738372		
OMG Oh My God! (2012)	61	3.942623		
Picnic (1955)	619	3.644588		
"Strawberry Statement	10	3.400000		
"Twilight Saga: Breaking Dawn - Part 2	1176	2.523384		
Shakespeare High (2011)	3	3.333333		
"Circle 2	1	1.750000		
Holiday Engagement (2011)	14	3.607143		
"Tiger Brigades	5	2.800000		
"Year One	6	3.083333		
"Big Blonde	2	3.000000		
Walter (1982)	2	4.000000		
10 Years (2011)	95	3.042105		
Price Check (2012)	24	2.937500		
Red Dawn (2012)	261	2.747126		
Starry starry night (Xing kong) (2011)	7	3.857143		
Rise of the Guardians (2012)	1527	3.526195		
"Seafarers	18	2.388889		
"Counterfeiters	6	3.666667		
Kabei: Our Mother (Kâbê) (2008)	5	4.200000		
Samurai Banners (Fürin kazan) (1969)	4	3.625000		
Murder on Flight 502 (1975)	2	3.000000		
Amor? (2011)	1	3.000000		
"Last Trapper	4	3.375000		
"Octopus	11	3.181818		
Locked Out (Enfermés dehors) (2006)	8	2.687500		
"Fantastic Fear of Everything	114	3.122807		
Fire Over England (1937)	4	2.750000		
Fog Over Frisco (1934)	4	2.625000		
Ten Seconds to Hell (1959)	3	3.000000		
Storage 24 (2012)	21	2.571429		
Tokio Baby (2010)	2	1.500000		
Deadfall (2012)	150	3.026667		
Skirt Day (La journée de la jupe) (2008)	10	3.650000		

acadgild@localhost:~ [delete hadoop log file... [acadgild@localhost:~]