

## **SESSION – 7**

### **CASE STUDY – MOVIE RATING ANALYSIS**

**Author:** Saravanan Ponnaiah

**Date:** 19-Feb-2019

### **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?

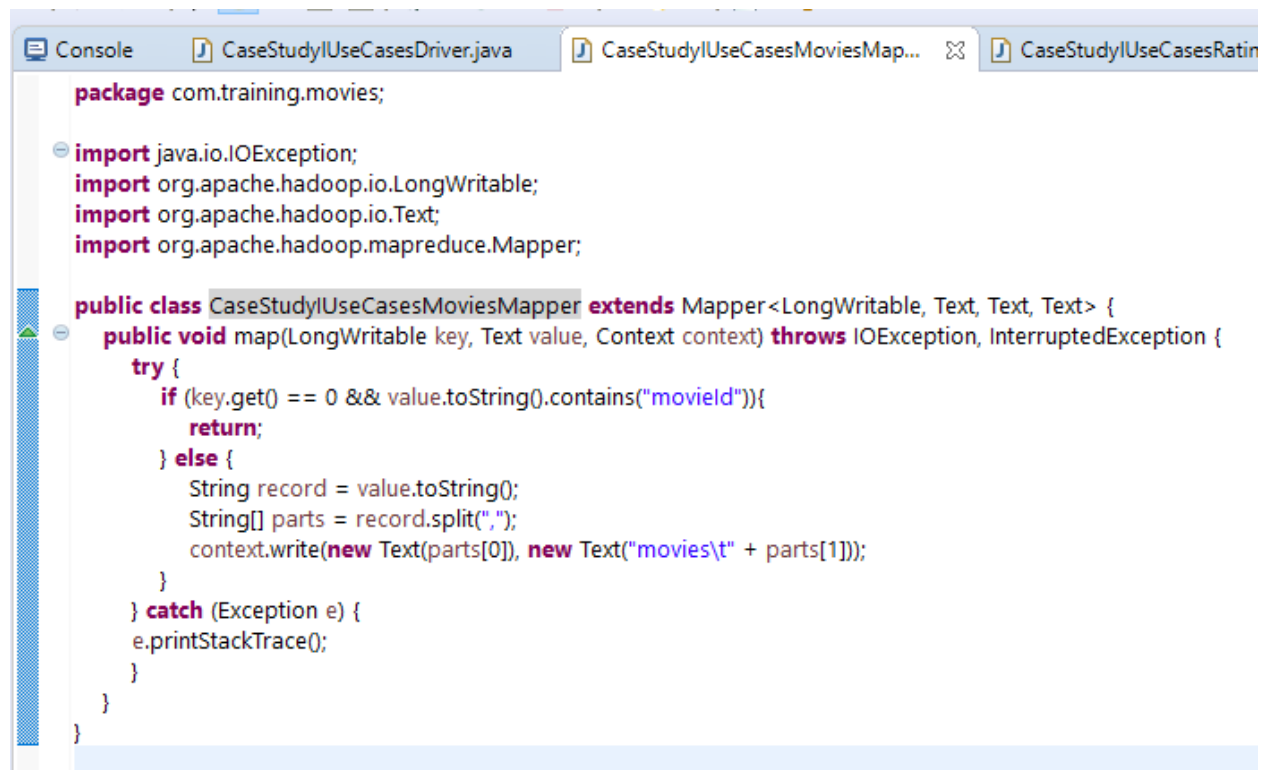
### **Input:**

- movies.csv
- ratings.csv

Both the input files are moved in to HDFS path “/mapreduce/movies”

### **Programs:**

CaseStudy\UseCases\MoviesMapper.java



```
package com.training.movies;

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

public class CaseStudy\UseCases\MoviesMapper 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();
        }
    }
}
```

## CaseStudyUseCasesRatingsMapper.java

```
package com.training.movies;

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

public class CaseStudyUseCasesRatingsMapper 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();
        }
    }
}
```

## CaseStudyUseCasesReducer.java

```
package com.training.movies;

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

public class CaseStudyUseCasesReducer 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));
    }
}
```

## CaseStudyIUseCasesDriver.java

```
package com.training.movies;

import org.apache.hadoop.conf.Configuration;

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);

        //job.setNumReduceTasks(0);

        //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);
    }
}
```

## Execution:

```
Applications  Places  System  acadgild@localhost: ~/install/hadoop/hadoop-2.6.5/bin  Tue Feb 19, 4
File Edit View Search Terminal Help
[acadgild@localhost bin]$ hadoop jar /home/acadgild/workspace/Mapreduce/MovieRatingAnalyzer.jar /map
reduce/movies/movies.csv /mapreduce/movies/ratings.csv /mapreduce/movies/out
```

## Execution Sequence:

```
acadgild@localhost:~/install/hadoop/hadoop-2.6.5/bin
File Edit View Search Terminal Help
19/02/19 20:33:06 INFO mapreduce.JobSubmitter: number of splits:7
19/02/19 20:33:06 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1550587858780_0001
19/02/19 20:33:07 INFO impl.YarnClientImpl: Submitted application application_1550587858780_0001
19/02/19 20:33:07 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/applicat
ion_1550587858780_0001/
19/02/19 20:33:07 INFO mapreduce.Job: Running job: job_1550587858780_0001
19/02/19 20:33:24 INFO mapreduce.Job: Job job_1550587858780_0001 running in uber mode : false
19/02/19 20:33:24 INFO mapreduce.Job: map 0% reduce 0%
19/02/19 20:34:26 INFO mapreduce.Job: map 1% reduce 0%
19/02/19 20:34:28 INFO mapreduce.Job: map 2% reduce 0%
19/02/19 20:34:29 INFO mapreduce.Job: map 3% reduce 0%
19/02/19 20:34:32 INFO mapreduce.Job: map 6% reduce 0%
19/02/19 20:34:35 INFO mapreduce.Job: map 7% reduce 0%
19/02/19 20:34:36 INFO mapreduce.Job: map 11% reduce 0%
19/02/19 20:34:39 INFO mapreduce.Job: map 14% reduce 0%
19/02/19 20:34:42 INFO mapreduce.Job: map 16% reduce 0%
19/02/19 20:34:43 INFO mapreduce.Job: map 17% reduce 0%
19/02/19 20:34:44 INFO mapreduce.Job: map 18% reduce 0%
19/02/19 20:34:46 INFO mapreduce.Job: map 19% reduce 0%
19/02/19 20:34:47 INFO mapreduce.Job: map 21% reduce 0%
19/02/19 20:34:49 INFO mapreduce.Job: map 22% reduce 0%
19/02/19 20:34:50 INFO mapreduce.Job: map 23% reduce 0%
19/02/19 20:34:51 INFO mapreduce.Job: map 24% reduce 0%
19/02/19 20:34:53 INFO mapreduce.Job: map 25% reduce 0%
19/02/19 20:34:54 INFO mapreduce.Job: map 26% reduce 0%
19/02/19 20:34:55 INFO mapreduce.Job: map 27% reduce 0%
19/02/19 20:34:57 INFO mapreduce.Job: map 28% reduce 0%
19/02/19 20:34:58 INFO mapreduce.Job: map 30% reduce 0%
19/02/19 20:35:00 INFO mapreduce.Job: map 32% reduce 0%
19/02/19 20:35:02 INFO mapreduce.Job: map 33% reduce 0%
19/02/19 20:35:03 INFO mapreduce.Job: map 34% reduce 0%
```

```
acadgild@localhost:~/install/hadoop/hadoop-2.6.5/bin
File Edit View Search Terminal Help
19/02/19 20:36:05 INFO mapreduce.Job: map 69% reduce 0%
19/02/19 20:36:06 INFO mapreduce.Job: map 70% reduce 0%
19/02/19 20:36:08 INFO mapreduce.Job: map 71% reduce 0%
19/02/19 20:36:09 INFO mapreduce.Job: map 72% reduce 0%
19/02/19 20:36:10 INFO mapreduce.Job: map 73% reduce 0%
19/02/19 20:36:13 INFO mapreduce.Job: map 75% reduce 0%
19/02/19 20:36:16 INFO mapreduce.Job: map 76% reduce 0%
19/02/19 20:36:55 INFO mapreduce.Job: map 76% reduce 10%
19/02/19 20:37:11 INFO mapreduce.Job: map 77% reduce 10%
19/02/19 20:37:17 INFO mapreduce.Job: map 78% reduce 10%
19/02/19 20:37:19 INFO mapreduce.Job: map 79% reduce 10%
19/02/19 20:37:21 INFO mapreduce.Job: map 80% reduce 10%
19/02/19 20:37:22 INFO mapreduce.Job: map 81% reduce 10%
19/02/19 20:37:24 INFO mapreduce.Job: map 82% reduce 10%
19/02/19 20:37:25 INFO mapreduce.Job: map 83% reduce 10%
19/02/19 20:37:27 INFO mapreduce.Job: map 84% reduce 10%
19/02/19 20:37:28 INFO mapreduce.Job: map 86% reduce 10%
19/02/19 20:37:30 INFO mapreduce.Job: map 87% reduce 10%
19/02/19 20:37:31 INFO mapreduce.Job: map 89% reduce 10%
19/02/19 20:37:33 INFO mapreduce.Job: map 90% reduce 10%
19/02/19 20:37:34 INFO mapreduce.Job: map 91% reduce 10%
19/02/19 20:37:37 INFO mapreduce.Job: map 93% reduce 10%
19/02/19 20:37:39 INFO mapreduce.Job: map 94% reduce 10%
19/02/19 20:37:40 INFO mapreduce.Job: map 96% reduce 10%
19/02/19 20:37:43 INFO mapreduce.Job: map 97% reduce 14%
19/02/19 20:37:44 INFO mapreduce.Job: map 98% reduce 14%
19/02/19 20:37:46 INFO mapreduce.Job: map 99% reduce 14%
19/02/19 20:37:50 INFO mapreduce.Job: map 100% reduce 24%
19/02/19 20:37:53 INFO mapreduce.Job: map 100% reduce 37%
19/02/19 20:37:56 INFO mapreduce.Job: map 100% reduce 67%
```

After mapper is completed and reducer is in progress, at some point, the execution awaits and does not proceed after sometime. When looking in to job history portal for that particular job, the message states that **"Waiting for AM container to be allocated, launched and registered with RM"**. Post that, the program does not proceed. I tried multiple times and same is the case,

Application application\_1550587858780\_0003 - Mozilla Firefox

Application application\_... x +

localhost:8088/cluster/app/application\_1550587858780\_0003


Search

☆

📁

⬇

🏠



Application

application\_1550587858780\_0003

Logged in as: i

Cluster

About

Nodes

Applications

NEW

NEW\_SAVING

SUBMITTED

ACCEPTED

RUNNING

FINISHED

FAILED

KILLED

Scheduler

Tools

Kill Application

Application Overview

User: acadgild

Name: CaseStudyIUseCase2Driver

Application Type: MAPREDUCE

Application Tags:

YarnApplicationState: ACCEPTED: waiting for AM container to be allocated, launched and register with RM.

FinalStatus Reported by AM: Application has not completed yet.

Started: 19-Feb-2019 21:48:33

Elapsed: 12mins, 4sec

Tracking URL: ApplicationMaster

Diagnostics:

Application Metrics

Total Resource Preempted: <memory:0, vCores:0>

Total Number of Non-AM Containers Preempted: 0

Total Number of AM Containers Preempted: 0

Resource Preempted from Current Attempt: <memory:0, vCores:0>

Number of Non-AM Containers Preempted from Current Attempt: 0

Aggregate Resource Allocation: 1825999 MB-seconds, 1508 vcore-seconds

Hence I could not validate the output of the program.