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"

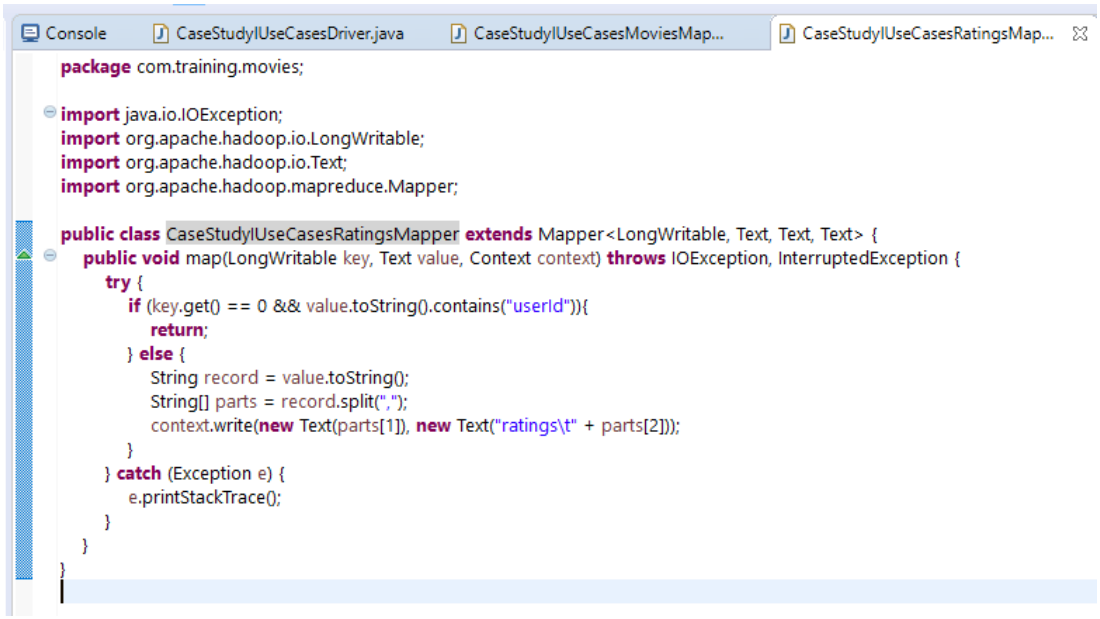
Programs:

CaseStudyIUseCasesMoviesMapper.java

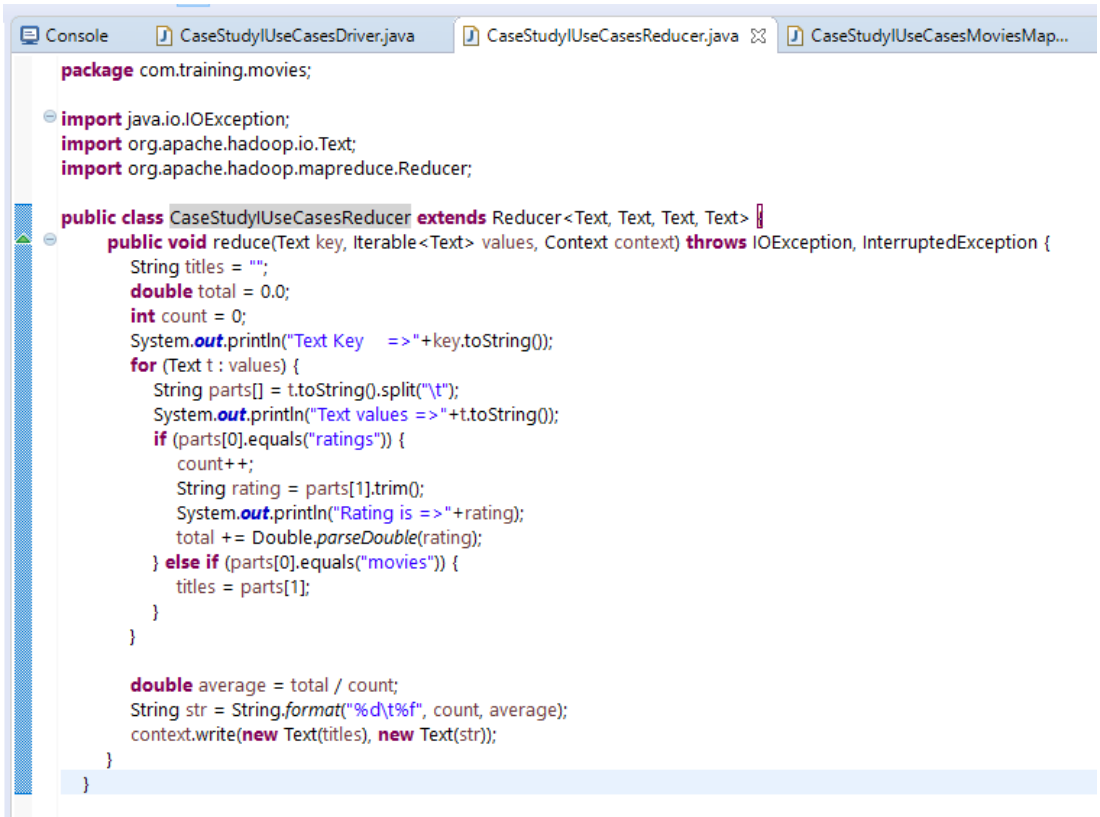
Console | CaseStudyIUseCasesDriver.java | CaseStudyIUseCasesMoviesMap... | CaseStudyIUseCasesRatin

```
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 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();
        }
    }
}
```

CaseStudyIUseCasesRatingsMapper.java



CaseStudyIUseCasesReducer.java



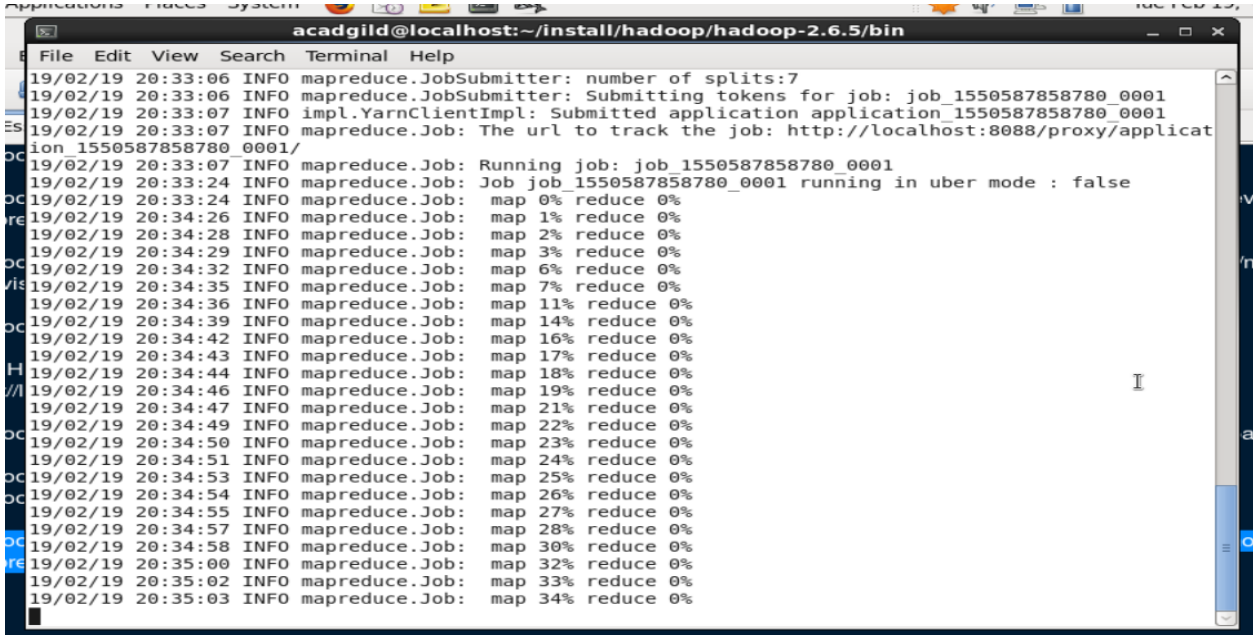
CaseStudyIUseCasesDriver.java

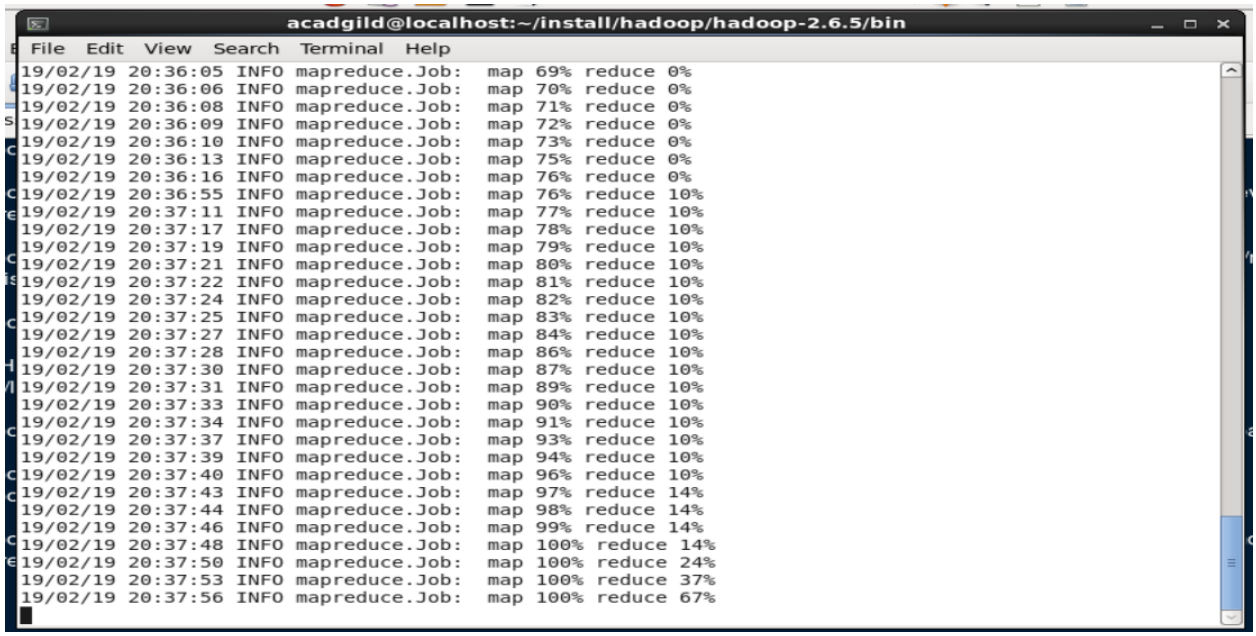
```java
package com.training.movies;

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:

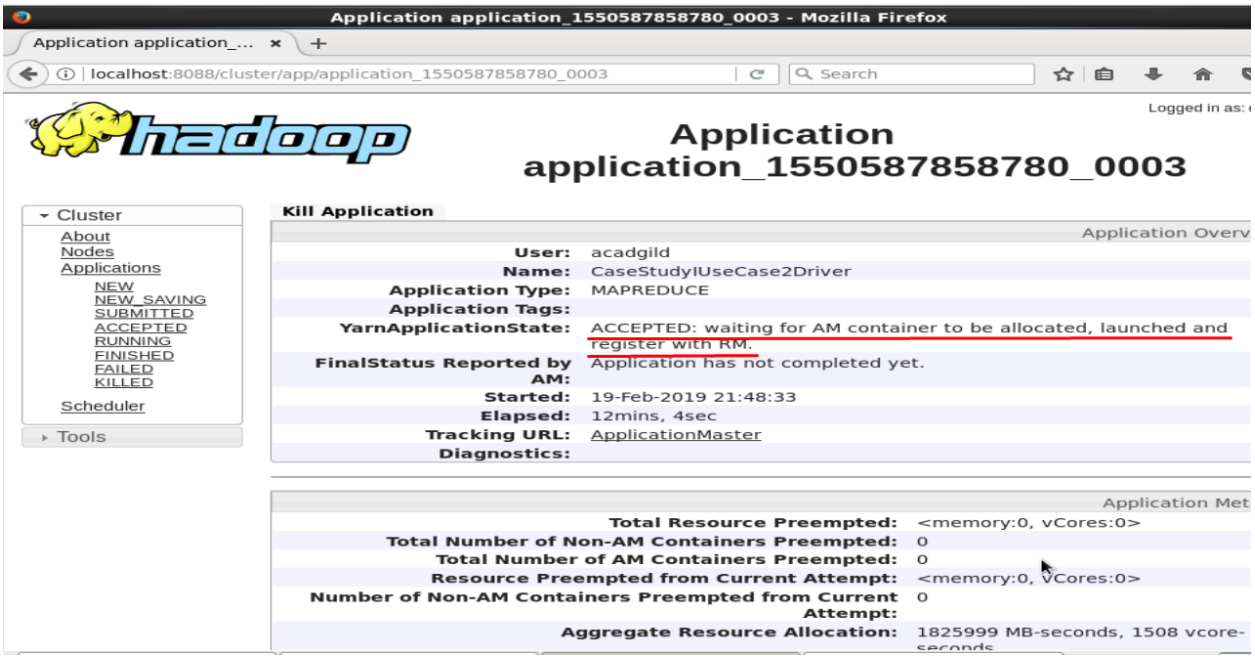
```
Applications  Places  System                                              Tue Feb 19,
acadgild@localhost:~/install/hadoop/hadoop-2.6.5/bin
File  Edit  View  Search  Terminal  Help
[acadgild@localhost bin]$ hadoop jar /home/acadgild/workspace/Mapreduce/MovieRatingAnalyzer.jar /map
reduce/movies/movies.csv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

Execution Sequence:





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,



Hence I could not validate the output of the program.