

Assignment 1: MapReduce Program Application

COS20028 – Big data Architecture and Applications

Semester 2, 2023

Name: [Nguyen Nam Tung](#)

Student ID: [103181157](#)

Date: [24/08/2023](#)

Driver Code

```
package stubs;

import org.apache.hadoop.fs.Path;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;
import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;
import org.apache.hadoop.mapreduce.Job;
import org.apache.hadoop.io.DoubleWritable;

/**
 * MapReduce jobs are typically implemented by using a driver class.
 * The purpose of a driver class is to set up the configuration for the
 * MapReduce job and to run the job.
 * Typical requirements for a driver class include configuring the input
 * and output data formats, configuring the map and reduce classes,
 * and specifying intermediate data formats.
 * |
 * The following is the code for the driver class:
 */
public class MovieGenreCount {

    public static void main(String[] args) throws Exception {

        /**
         * The expected command-line arguments are the paths containing
         * input and output data. Terminate the job if the number of
         * command-line arguments is not exactly 2.
         */
        if (args.length != 2) {
            System.out.printf(
                "Usage: WordCount <input dir> <output dir>\n");
            System.exit(-1);
        }
    }
}
```

```

    * Specify the paths to the input and output data based on the
    * command-line arguments.
    */
FileInputFormat.setInputPaths(job, new Path(args[0]));
FileOutputFormat.setOutputPath(job, new Path(args[1]));

/*
 * Specify the mapper and reducer classes.
 */
job.setMapperClass(MovieGenreMapper.class);
job.setReducerClass(MovieGenreReducer.class);

System.out.print("Programmer: Nguyen Nam Tung\nStudent ID: 103181157\n");

job.setOutputKeyClass(Text.class);
job.setOutputValueClass(DoubleWritable.class);

System.out.println("Programmer: Tung");
System.out.println("Student ID: 103181157");

job.setMapOutputKeyClass(Text.class);
job.setMapOutputValueClass(DoubleWritable.class);

/*
 * Start the MapReduce job and wait for it to finish.
 * If it finishes successfully, return 0. If not, return 1.
 */
boolean success = job.waitForCompletion(true);
System.exit(success ? 0 : 1);
}
}

```

Mapper Code

```

package stubs;

import java.io.IOException;

public class MovieGenreMapper extends Mapper<LongWritable, Text, Text, DoubleWritable> {

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

        String[] record = value.toString().split(",");
        //split the record value by ,
        if (record[0].equals("Colour") || record[0].equals(" Black and White")){
            // if the type is color or black and white
            double Likesnumber = Double.parseDouble(record[record.length - 1]);
            //get the like number and convert it to double
            context.write(new Text(record[0]), new DoubleWritable(Likesnumber));
        }
    }
}

```

Reducer Code

```

package stubs;

import java.io.IOException;

public class MovieGenreReducer extends Reducer<Text, DoubleWritable, Text, DoubleWritable> {

    @Override
    public void reduce(Text key, Iterable<DoubleWritable> values, Context context)
        throws IOException, InterruptedException {
        int count = 0;
        int sum = 0;
        for (DoubleWritable value : values) {
            // loops through every values of the likes number
            count++;
            sum += value.get();
        }
        double average = (double)sum/(double)count;
        //get the average by taking the sum of the movie / number of movies
        context.write(key, new DoubleWritable(average));
    }
}

```

Commands and the Execution Screenshot

```

[training@localhost ~]$ cd workspace/moviecount/src
[training@localhost src]$ javac -classpath `hadoop classpath` stubs/*.java
[training@localhost src]$ jar cvf wc.jar stubs/*.class
added manifest
adding: stubs/MovieGenreCount.class(in = 1840) (out= 1002)(deflated 45%)
adding: stubs/MovieGenreMapper.class(in = 1804) (out= 739)(deflated 59%)
adding: stubs/MovieGenreReducer.class(in = 1654) (out= 699)(deflated 57%)

[training@localhost src]$ hadoop jar wc.jar stubs.MovieGenreCount movie_metadata
.csv movie_res1
Programmer: Nguyen Nam Tung
Student ID: 103181157
Programmer: Tung
Student ID: 103181157
23/08/24 22:46:52 WARN mapred.JobClient: Use GenericOptionsParser for parsing the
arguments. Applications should implement Tool for the same.
23/08/24 22:46:52 INFO input.FileInputFormat: Total input paths to process : 1

[training@localhost ~]$ hdfs dfs -get movie_res1
get: `movie_res1/_SUCCESS`: File exists
get: `movie_res1/_logs/history/job_202308241833_0009_1692931612586_training_Word
Count`: File exists
get: `movie_res1/_logs/history/job_202308241833_0009_conf.xml`: File exists
get: `movie_res1/part-r-000000`: File exists
[training@localhost ~]$ less movie_res1/part-r-000000

```

First Page of the Output File Screenshot

| | | | | | | | | | |
|--------|-----------|-----------|-------------------|---------|----------|----------|--------------------|---------|--|
| Action | 4398.2 | | | | | | | | |
| Action | Adventure | | 7769.727272727273 | | | | | | |
| Action | Adventure | Animation | Comedy | Crime | Family | Fantasy | 2000.0 | | |
| Action | Adventure | Animation | Comedy | Drama | Family | Sci-Fi | 20500.0 | | |
| Action | Adventure | Animation | Comedy | Family | | | 8368.5 | | |
| Action | Adventure | Animation | Comedy | Family | Fantasy | | 31500.0 | | |
| Action | Adventure | Animation | Comedy | Family | Fantasy | Sci-Fi | 0.0 | | |
| Action | Adventure | Animation | Comedy | Family | Sci-Fi | | 141.66666666666666 | | |
| Action | Adventure | Animation | Comedy | Fantasy | | | 408.0 | | |
| Action | Adventure | Animation | Comedy | Fantasy | Sci-Fi | | 307.0 | | |
| Action | Adventure | Animation | Comedy | Sci-Fi | | | 77.0 | | |
| Action | Adventure | Animation | Drama | Fantasy | Sci-Fi | | 2000.0 | | |
| Action | Adventure | Animation | Drama | Mystery | Sci-Fi | Thriller | | 13000.0 | |
| Action | Adventure | Animation | Family | | | | 0.0 | | |
| Action | Adventure | Animation | Family | Fantasy | | | 8062.0 | | |
| Action | Adventure | Animation | Family | Fantasy | Sci-Fi | | 19.0 | | |
| Action | Adventure | Animation | Family | Sci-Fi | | | 103.5 | | |
| Action | Adventure | Animation | Family | Sci-Fi | Thriller | | 973.0 | | |
| Action | Adventure | Animation | Fantasy | | | | 3000.0 | | |
| Action | Adventure | Animation | Fantasy | Romance | Sci-Fi | | 0.0 | | |
| Action | Adventure | Biography | Drama | History | | | 31000.0 | | |
| Action | Adventure | Biography | Drama | History | Romance | War | 0.0 | | |
| Action | Adventure | Biography | Drama | History | Thriller | | 27000.0 | | |