

Jaeseok Huh  
2015005241  
14 November 2018

# Distributed Computing

## Assignment #2

### 1. Directory Structure

averageRating.jar  
build.sh\*  
com/jaeseok/  
    *compiled classes*  
data/  
    *ratings.csv*  
doc  
.gitignore  
run.sh\*  
src/  
    AverageRating.java - includes MapReducer  
    PairWritable.java - Pair<Double, Int>

### 2. A Brief Explanation about Code

In AverageRating.java, there are two inner classes, “TokenizerMapper”(line 24) and “AverageReducer”(line 54). They serve as a wrapper class for mapper and reducer respectively.

In “TokenizerMapper”, `public void map(Object, Text, Context)` (line serves as a mapper function and later added to the job(line 75). As stated in the documentation, it maps the data into Pair<Double, Int>, denoting a rating of the user and a count (one) respectively.

Since Hadoop requires to access the fields, class Pair and its member shall be “WriteComparable” (See PairWritable.java)

In “AverageReducer”, it provides reducer(line 58). It iterates through the intermediate lists and get the average value of them. It finally writes to a context(line 67).

In java main, a job hadoop shall execute is specified. Since the output values of Mapper and Reducer have different type, you should use job.setMapOutputValueClass() to clarify it.

### 3. How to Build & Run

./build.sh && ./run.sh

### 4. Screenshot

```
root@ip-172-31-29-219:~/Distributed-System/project2# ./build.sh && ./run.sh
    mkdir: '/input': File exists
copyFromLocal: '/input/ratings.csv': File exists
18/11/14 12:15:20 INFO fs.TrashPolicyDefault: Namenode trash configuration: Deletion interval = 0 minutes, Emptier interval = 0 minutes.
Deleted /output
18/11/14 12:15:22 INFO client.RMProxy: Connecting to ResourceManager at Master/172.31.29.219:8032
18/11/14 12:15:23 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to
remedy this.
18/11/14 12:15:23 INFO input.FileInputFormat: Total input paths to process : 1
18/11/14 12:15:23 INFO mapreduce.JobSubmitter: number of splits:1
18/11/14 12:15:24 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1542173541650_0028
18/11/14 12:15:24 INFO impl.YarnClientImpl: Submitted application application_1542173541650_0028
18/11/14 12:15:24 INFO mapreduce.Job: The url to track the job: http://master:8088/proxy/application_1542173541650_0028/
18/11/14 12:15:24 INFO mapreduce.Job: Running job: job_1542173541650_0028
18/11/14 12:15:32 INFO mapreduce.Job: Job job_1542173541650_0028 running in uber mode : false
18/11/14 12:15:32 INFO mapreduce.Job:  map 0% reduce 0%
18/11/14 12:15:39 INFO mapreduce.Job:  map 100% reduce 0%
18/11/14 12:15:51 INFO mapreduce.Job:  map 100% reduce 50%
18/11/14 12:15:52 INFO mapreduce.Job:  map 100% reduce 100%
18/11/14 12:15:53 INFO mapreduce.Job: Job job_1542173541650_0028 completed successfully
18/11/14 12:15:53 INFO mapreduce.Job: Counters: 50
    File System Counters
        FILE: Number of bytes read=1798641
        FILE: Number of bytes written=3918808
        FILE: Number of read operations=0
        FILE: Number of large read operations=0
        FILE: Number of write operations=0
        HDFS: Number of bytes read=2483824
        HDFS: Number of bytes written=12396
        HDFS: Number of read operations=9
        HDFS: Number of large read operations=0
        HDFS: Number of write operations=4
    Job Counters
        Killed reduce tasks=1
        Launched map tasks=1
        Launched reduce tasks=2
        Data-local map tasks=1
        Total time spent by all maps in occupied slots (ms)=5411
        Total time spent by all reduces in occupied slots (ms)=19027
        Total time spent by all map tasks (ms)=5411
        Total time spent by all reduce tasks (ms)=19027
        Total vcore-milliseconds taken by all map tasks=5411
        Total vcore-milliseconds taken by all reduce tasks=19027
        Total megabyte-milliseconds taken by all map tasks=5540864
        Total megabyte-milliseconds taken by all reduce tasks=19483648
    Map-Reduce Framework
        Map input records=100837
        Map output records=100836
        Map output bytes=1596957
        Map output materialized bytes=1798641
        Input split bytes=101
        Combine input records=0
        Combine output records=0
        Reduce input groups=610
        Reduce shuffle bytes=1798641
        Reduce input records=100836
        Reduce output records=610
        Spilled Records=201672
        Shuffled Maps =2
        Failed Shuffles=0
        Merged Map outputs=2
        GC time elapsed (ms)=208
        CPU time spent (ms)=3860
        Physical memory (bytes) snapshot=429408256
        Virtual memory (bytes) snapshot=2389426176
        Total committed heap usage (bytes)=200282112
    Shuffle Errors
        BAD_ID=0
        CONNECTION=0
        IO_ERROR=0
        WRONG_LENGTH=0
        WRONG_MAP=0
        WRONG_REDUCE=0
    File Input Format Counters
        Bytes Read=2483723
    File Output Format Counters
        Bytes Written=12396
root@ip-172-31-29-219:~/Distributed-System/project2#
```

```

root@ip-172-31-29-219:~/Distributed-System/project2# hdfs dfs -text /output^C
root@ip-172-31-29-219:~/Distributed-System/project2# hdfs dfs -ls /output
Found 3 items
-rw-r--r--  2 root supergroup          0 2018-11-14 12:15 /output/_SUCCESS
-rw-r--r--  2 root supergroup      6408 2018-11-14 12:15 /output/part-r-00000
-rw-r--r--  2 root supergroup      5988 2018-11-14 12:15 /output/part-r-00001
root@ip-172-31-29-219:~/Distributed-System/project2# hdfs dfs -text /output/_SUCCESS
root@ip-172-31-29-219:~/Distributed-System/project2# hdfs dfs -text /output/part-r-00000
1      4.366379310344827
10     3.2785714285714285
100    3.945945945945946
102    3.357142857142857
104    3.5073260073260073
106    4.4393939393939394
108    3.986842105263158
111    3.3397832817337463
113    3.6466666666666665
115    3.767857142857143
117    3.3393939393939394
119    4.176744186046512
12     4.390625
120    3.409090909090909
122    4.546232876712328
124    3.99
126    3.289473684210526
128    4.363636363636363
131    3.4420289855072466

```