Solution Assignment 3

wordmean: A map/reduce program that counts the average length of the words in the input files.

```
[acadgild@localhost mapreduce]$ hadoop jar hadoop-mapreduce-examples-2.9.0.jar wordmean /hadoopdata/hdfs/file1.txt /hadoopdata/hdfs/out/wordmeanfile1
18/01/23 19:16:42 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/01/23 19:16:43 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/01/23 19:16:45 INFO input.FileInputFormat: Total input files to process : 1
18/01/23 19:16:45 INFO mapreduce.JobSubmitter: number of splits:1
18/01/23 19:16:45 INFO Configuration.deprecation: yarn.resourcemanager.system-metrics-publisher.enabled is deprecated. Instead, use yarn.system-metrics-publisher.ena
d
18/01/23 19:16:45 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1516707793508_0012
18/01/23 19:16:46 INFO impl.YarnClientImpl: Submitted application application_1516707793508_0012
18/01/23 19:16:46 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1516707793508_0012/
18/01/23 19:16:46 INFO mapreduce.Job: Running job: job_1516707793508_0012
18/01/23 19:16:56 INFO mapreduce.Job: Job job_1516707793508_0012 running in uber mode : false
18/01/23 19:16:56 INFO mapreduce.Job:  map 0% reduce 0%
18/01/23 19:17:02 INFO mapreduce.Job:  map 100% reduce 0%
18/01/23 19:17:10 INFO mapreduce.Job:  map 100% reduce 100%
18/01/23 19:17:10 INFO mapreduce.Job: Job job_1516707793508_0012 completed successfully
18/01/23 19:17:10 INFO mapreduce.Job: Counters: 49
        File System Counters
                FILE: Number of bytes read=39
                FILE: Number of bytes written=404383
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=162
                HDFS: Number of bytes written=18
                HDFS: Number of read operations=6
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=2
        Job Counters
                Launched map tasks=1
                Launched reduce tasks=1
                Data-local map tasks=1
                Total time spent by all maps in occupied slots (ms)=4134
                Total time spent by all reduces in occupied slots (ms)=4761
                Total time spent by all map tasks (ms)=4134
                Total time spent by all reduce tasks (ms)=4761
                Total vcore-milliseconds taken by all map tasks=4134
                Total vcore-milliseconds taken by all reduce tasks=4761
                Total megabyte-milliseconds taken by all map tasks=4233216
                Total megabyte-milliseconds taken by all reduce tasks=4875264
        Map-Reduce Framework
```

```
                Map input records=1
                Map output records=16
                Map output bytes=232
                Map output materialized bytes=39
                Input split bytes=112
                Combine input records=16
                Combine output records=2
                Reduce input groups=2
                Reduce shuffle bytes=39
                Reduce input records=2
                Reduce output records=2
                Spilled Records=4
                Shuffled Maps =1
                Failed Shuffles=0
                Merged Map outputs=1
                GC time elapsed (ms)=143
                CPU time spent (ms)=1210
                Physical memory (bytes) snapshot=346324992
                Virtual memory (bytes) snapshot=4127256576
                Total committed heap usage (bytes)=222429184
        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=50
        File Output Format Counters
                Bytes Written=18
The mean is: 5.25
[acadgild@localhost mapreduce]$
```

wordmedian: A map/reduce program that counts the median length of the words in the input files.

```
[acadgild@localhost mapreduce]$ hadoop jar hadoop-mapreduce-examples-2.9.0.jar wordmedian /hadoopdata/hdfs/file1.txt /hadoopdata/hdfs/out/wordmedianfile1
18/01/23 19:19:25 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/01/23 19:19:26 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/01/23 19:19:27 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with
oolRunner to remedy this.
18/01/23 19:19:27 INFO input.FileInputFormat: Total input files to process : 1
18/01/23 19:19:27 INFO mapreduce.JobSubmitter: number of splits:1
18/01/23 19:19:27 INFO Configuration.deprecation: yarn.resourcemanager.system-metrics-publisher.enabled is deprecated. Instead, use yarn.system-metrics-publisher.enab
d
18/01/23 19:19:28 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1516707793508_0013
18/01/23 19:19:28 INFO impl.YarnClientImpl: Submitted application application_1516707793508_0013
18/01/23 19:19:28 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1516707793508_0013/
18/01/23 19:19:28 INFO mapreduce.Job: Running job: job_1516707793508_0013
18/01/23 19:19:38 INFO mapreduce.Job: Job job_1516707793508_0013 running in uber mode : false
18/01/23 19:19:38 INFO mapreduce.Job:  map 0% reduce 0%
18/01/23 19:19:44 INFO mapreduce.Job:  map 100% reduce 0%
18/01/23 19:19:52 INFO mapreduce.Job:  map 100% reduce 100%
18/01/23 19:19:52 INFO mapreduce.Job: Job job_1516707793508_0013 completed successfully
18/01/23 19:19:52 INFO mapreduce.Job: Counters: 49
        File System Counters
                FILE: Number of bytes read=56
                FILE: Number of bytes written=404117
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=162
                HDFS: Number of bytes written=20
                HDFS: Number of read operations=6
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=2
        Job Counters
                Launched map tasks=1
                Launched reduce tasks=1
                Data-local map tasks=1
                Total time spent by all maps in occupied slots (ms)=4008
                Total time spent by all reduces in occupied slots (ms)=4427
                Total time spent by all map tasks (ms)=4008
                Total time spent by all reduce tasks (ms)=4427
                Total vcore-milliseconds taken by all map tasks=4008
                Total vcore-milliseconds taken by all reduce tasks=4427
                Total megabyte-milliseconds taken by all map tasks=4104192
                Total megabyte-milliseconds taken by all reduce tasks=4533248
        Map-Reduce Framework
                Map input records=1
                Map output records=8
                Map output bytes=64
                Map output materialized bytes=56
                Input split bytes=112
                Combine input records=8
                Combine output records=5
                Reduce input groups=5
                Reduce shuffle bytes=56
                Reduce input records=5
                Reduce output records=5
                Spilled Records=10
                Shuffled Maps =1
                Failed Shuffles=0
                Merged Map outputs=1
                GC time elapsed (ms)=129
                CPU time spent (ms)=1170
                Physical memory (bytes) snapshot=346796032
                Virtual memory (bytes) snapshot=4127256576
                Total committed heap usage (bytes)=222429184
        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=50
        File Output Format Counters
                Bytes Written=20
The median is: 5
[acadgild@localhost mapreduce]$
```

wordstandarddeviation: A map/reduce program that counts the standard deviation of the length of the words in the input files.

```
[acadgild@localhost mapreduce]$ hadoop jar hadoop-mapreduce-examples-2.9.0.jar wordstandarddeviation /hadoopdata/hdfs/file1.txt /hadoopdata/hdfs/out/wordstandarddeviati
onfile1
18/01/23 19:22:22 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/01/23 19:22:23 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/01/23 19:22:24 INFO input.FileInputFormat: Total input files to process : 1
18/01/23 19:22:24 WARN hdfs.DataStreamer: Caught exception
java.lang.InterruptedException
        at java.lang.Object.wait(Native Method)
        at java.lang.Thread.join(Thread.java:1252)
        at java.lang.Thread.join(Thread.java:1326)
        at org.apache.hadoop.hdfs.DataStreamer.closeResponder(DataStreamer.java:980)
        at org.apache.hadoop.hdfs.DataStreamer.endBlock(DataStreamer.java:630)
        at org.apache.hadoop.hdfs.DataStreamer.run(DataStreamer.java:807)
18/01/23 19:22:24 INFO mapreduce.JobSubmitter: number of splits:1
18/01/23 19:22:25 INFO Configuration.deprecation: yarn.resourcemanager.system-metrics-publisher.enabled is deprecated. Instead, use yarn.system-metrics-publisher.enable
d
18/01/23 19:22:25 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1516707793508_0014
18/01/23 19:22:25 INFO impl.YarnClientImpl: Submitted application application_1516707793508_0014
18/01/23 19:22:25 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1516707793508_0014/
18/01/23 19:22:25 INFO mapreduce.Job: Running job: job_1516707793508_0014
18/01/23 19:22:35 INFO mapreduce.Job: Job job_1516707793508_0014 running in uber mode : false
18/01/23 19:22:35 INFO mapreduce.Job:  map 0% reduce 0%
18/01/23 19:22:42 INFO mapreduce.Job:  map 100% reduce 0%
18/01/23 19:22:49 INFO mapreduce.Job:  map 100% reduce 100%
18/01/23 19:22:50 INFO mapreduce.Job: Job job_1516707793508_0014 completed successfully
18/01/23 19:22:50 INFO mapreduce.Job: Counters: 49
        File System Counters
                FILE: Number of bytes read=56
                FILE: Number of bytes written=404603
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=162
                HDFS: Number of bytes written=29
                HDFS: Number of read operations=6
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=2
        Job Counters
                Launched map tasks=1
                Launched reduce tasks=1
                Data-local map tasks=1
                Total time spent by all maps in occupied slots (ms)=4086
                Total time spent by all reduces in occupied slots (ms)=4728
```

```
                Total time spent by all map tasks (ms)=4086
                Total time spent by all reduce tasks (ms)=4728
                Total vcore-milliseconds taken by all map tasks=4086
                Total vcore-milliseconds taken by all reduce tasks=4728
                Total megabyte-milliseconds taken by all map tasks=4184064
                Total megabyte-milliseconds taken by all reduce tasks=4841472
        Map-Reduce Framework
                Map input records=1
                Map output records=24
                Map output bytes=352
                Map output materialized bytes=56
                Input split bytes=112
                Combine input records=24
                Combine output records=3
                Reduce input groups=3
                Reduce shuffle bytes=56
                Reduce input records=3
                Reduce output records=3
                Spilled Records=6
                Shuffled Maps =1
                Failed Shuffles=0
                Merged Map outputs=1
                GC time elapsed (ms)=139
                CPU time spent (ms)=1330
                Physical memory (bytes) snapshot=345300992
                Virtual memory (bytes) snapshot=4127256576
                Total committed heap usage (bytes)=222429184
        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=50
        File Output Format Counters
                Bytes Written=29
The standard deviation is: 2.436698586202241
[acadgild@localhost mapreduce]$
```