

ASSIGNMENT 3.1

Task: To execute the **WordMedian**, **WordMean** and **WordStandardDeviation** programs using **hadoop-mapreduce-examples-2.9.0.jar**.

Input File: File1.txt

acadgild@localhost:~

```
[acadgild@localhost ~]$ hdfs dfs -cat file1.txt
18/01/28 15:30:47 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... u
A B C
C A B
B B B
B C A
D D D
D D D
D A C
D D B
[acadgild@localhost ~]$
```

Command to execute WordMean: `hadoop jar /home/acadgild/Desktop/hadoop-mapreduce-examples-2.9.0.jar wordmean file1.txt wordmean`

acadgild@localhost:~

```
[acadgild@localhost ~]$
[acadgild@localhost ~]$ hadoop jar /home/acadgild/Desktop/hadoop-mapreduce-examples-2.9.0.jar wordmean file1.txt wordmean
18/01/28 15:35:23 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where
18/01/28 15:35:24 INFO client.RMProxy: Connecting to Resource Manager at localhost/127.0.0.1:8032
18/01/28 15:35:26 INFO input.FileInputFormat: Total input files to process : 1
18/01/28 15:35:26 WARN hdfs.DataStreamer: Caught exception
java.lang.InterruptedExce
    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)
18/01/28 15:35:26 WARN hdfs.DataStreamer: Caught exception
java.lang.InterruptedExce
    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)
18/01/28 15:35:26 INFO mapreduce.JobSubmitter: number of splits:1
18/01/28 15:35:26 INFO Configuration.deprecation: yarn.resourcemanager.system-metrics-publisher.enabled is deprecated. Instead, use yarn
d
18/01/28 15:35:27 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1517054117288_0022
18/01/28 15:35:27 INFO impl.YarnClientImpl: Submitted application application_1517054117288_0022
18/01/28 15:35:27 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1517054117288_0022/
18/01/28 15:35:27 INFO mapreduce.Job: Running job: job_1517054117288_0022
18/01/28 15:35:41 INFO mapreduce.Job: Job job_1517054117288_0022 running in uber mode : false
18/01/28 15:35:41 INFO mapreduce.Job: map 0% reduce 0%
18/01/28 15:35:48 INFO mapreduce.Job: map 100% reduce 0%
18/01/28 15:35:56 INFO mapreduce.Job: map 100% reduce 100%
18/01/28 15:35:57 INFO mapreduce.Job: Job job_1517054117288_0022 completed successfully
18/01/28 15:35:57 INFO mapreduce.Job: Counters: 49
File System Counters
    FILE: Number of bytes read=39
    FILE: Number of bytes written=404357
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=158
    HDFS: Number of bytes written=19
    HDFS: Number of read operations=6
```

acadmild@localhost:~

```
Launched reduce tasks=1
Data-local map tasks=1
Total time spent by all maps in occupied slots (ms)=5161
Total time spent by all reduces in occupied slots (ms)=5684
Total time spent by all map tasks (ms)=5161
Total time spent by all reduce tasks (ms)=5684
Total vcore-milliseconds taken by all map tasks=5161
Total vcore-milliseconds taken by all reduce tasks=5684
Total megabyte-milliseconds taken by all map tasks=5284864
Total megabyte-milliseconds taken by all reduce tasks=5820416
Map-Reduce Framework
  Map input records=8
  Map output records=48
  Map output bytes=696
  Map output materialized bytes=39
  Input split bytes=110
  Combine input records=48
  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)=150
  CPU time spent (ms)=1490
  Physical memory (bytes) snapshot=349519872
  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=48
File Output Format Counters
  Bytes Written=19
The mean is: 1.0
```

Output

acadmild@localhost:~

```
[acadmild@localhost ~]$ hdfs dfs -cat wordmedian/part-r-00000
18/01/28 15:42:58 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... us
count  24
length 24
[acadmild@localhost ~]$
```

Command to execute WordMedian: `hadoop jar /home/acadmild/Desktop/hadoop-mapreduce-examples-2.9.0.jar wordmedian file1.txt wordmedian`

acadmild@localhost:~

```
[acadmild@localhost ~]$
[acadmild@localhost ~]$ hadoop jar /home/acadmild/Desktop/hadoop-mapreduce-examples-2.9.0.jar wordmedian file1.txt wordmedian
18/01/28 15:44:28 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes w
18/01/28 15:44:30 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/01/28 15:44:31 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface
oolRunner to remedy this.
18/01/28 15:44:31 INFO input.FileInputFormat: Total input files to process : 1
18/01/28 15:44:31 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)
18/01/28 15:44:32 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)
18/01/28 15:44:32 INFO mapreduce.JobSubmitter: number of splits:1
18/01/28 15:44:32 INFO Configuration.deprecation: yarn.resourcemanager.system-metrics-publisher.enabled is deprecated. Instead, use
d
18/01/28 15:44:32 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1517054117288_0023
18/01/28 15:44:33 INFO impl.YarnClientImpl: Submitted application application_1517054117288_0023
18/01/28 15:44:33 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1517054117288_0023/
18/01/28 15:44:33 INFO mapreduce.Job: Running job: job_1517054117288_0023
18/01/28 15:44:46 INFO mapreduce.Job: Job job_1517054117288_0023 running in uber mode : false
18/01/28 15:44:46 INFO mapreduce.Job:  map 0% reduce 0%
18/01/28 15:44:53 INFO mapreduce.Job:  map 100% reduce 0%
18/01/28 15:45:02 INFO mapreduce.Job:  map 100% reduce 100%
18/01/28 15:45:02 INFO mapreduce.Job: Job job_1517054117288_0023 completed successfully
18/01/28 15:45:03 INFO mapreduce.Job: Counters: 49
    File System Counters
        FILE: Number of bytes read=16
        FILE: Number of bytes written=404011
        FILE: Number of read operations=0
        FILE: Number of large read operations=0
        FILE: Number of write operations=0
        HDFS: Number of bytes read=158
```

acadmild@localhost:~

```
Launched reduce tasks=1
Data-local map tasks=1
Total time spent by all maps in occupied slots (ms)=5387
Total time spent by all reduces in occupied slots (ms)=6183
Total time spent by all map tasks (ms)=5387
Total time spent by all reduce tasks (ms)=6183
Total vcore-milliseconds taken by all map tasks=5387
Total vcore-milliseconds taken by all reduce tasks=6183
Total megabyte-milliseconds taken by all map tasks=5516288
Total megabyte-milliseconds taken by all reduce tasks=6331392
Map-Reduce Framework
  Map input records=8
  Map output records=24
  Map output bytes=192
  Map output materialized bytes=16
  Input split bytes=110
  Combine input records=24
  Combine output records=1
  Reduce input groups=1
  Reduce shuffle bytes=16
  Reduce input records=1
  Reduce output records=1
  Spilled Records=2
  Shuffled Maps =1
  Failed Shuffles=0
  Merged Map outputs=1
  GC time elapsed (ms)=182
  CPU time spent (ms)=1560
  Physical memory (bytes) snapshot=348508160
  Virtual memory (bytes) snapshot=4127416320
  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=48
File Output Format Counters
  Bytes Written=5
The median is: 1
```

Output

acadmild@localhost:~

```
[acadmild@localhost ~]$ hdfs dfs -cat wordmedian/part-r-00000
18/01/28 15:54:59 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform...
1      24
[acadmild@localhost ~]$
```

Command to execute WordStandardDeviation :

hadoop jar /home/acadgild/Desktop/hadoop-mapreduce-examples-2.9.0.jar
wordstandarddeviation file1.txt WordStandardDeviation

```
acadgild@localhost:~$  
[acadgild@localhost ~]$  
[acadgild@localhost ~]$ hadoop jar /home/acadgild/Desktop/hadoop-mapreduce-examples-2.9.0.jar wordstandarddeviation file1.txt WordStandardDeviation  
18/01/28 16:04:52 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable  
18/01/28 16:04:53 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032  
18/01/28 16:04:55 INFO input.FileInputFormat: Total input files to process : 1  
18/01/28 16:04:55 WARN hdfs.DataStreamer: Caught exception  
java.lang.InterruptedExecution  
    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)  
18/01/28 16:04:55 WARN hdfs.DataStreamer: Caught exception  
java.lang.InterruptedExecution  
    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)  
18/01/28 16:04:55 INFO mapreduce.JobSubmitter: number of splits:1  
18/01/28 16:04:56 INFO Configuration.deprecation: yarn.resourcemanager.system-metrics-publisher.enabled is deprecated. Instead, use yarn.system-metric  
d  
18/01/28 16:04:56 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1517054117288_0025  
18/01/28 16:04:56 INFO impl.YarnClientImpl: Submitted application application_1517054117288_0025  
18/01/28 16:04:57 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1517054117288_0025/  
18/01/28 16:04:57 INFO mapreduce.Job: Running job: job_1517054117288_0025  
18/01/28 16:05:10 INFO mapreduce.Job: Job job_1517054117288_0025 running in uber mode : false  
18/01/28 16:05:10 INFO mapreduce.Job:  map 0% reduce 0%  
18/01/28 16:05:17 INFO mapreduce.Job:  map 100% reduce 0%  
18/01/28 16:05:25 INFO mapreduce.Job:  map 100% reduce 100%  
18/01/28 16:05:25 INFO mapreduce.Job: Job job_1517054117288_0025 completed successfully  
18/01/28 16:05:26 INFO mapreduce.Job: Counters: 49  
    File System Counters  
        FILE: Number of bytes read=56  
        FILE: Number of bytes written=404577  
        FILE: Number of read operations=0  
        FILE: Number of large read operations=0  
        FILE: Number of write operations=0  
        HDFS: Number of bytes read=158  
        HDFS: Number of bytes written=29  
        HDFS: Number of read operations=6  
        HDFS: Number of large read operations=0
```

acadgild@localhost:~

```
Launched map tasks=1
Launched reduce tasks=1
Data-local map tasks=1
Total time spent by all maps in occupied slots (ms)=5326
Total time spent by all reduces in occupied slots (ms)=5773
Total time spent by all map tasks (ms)=5326
Total time spent by all reduce tasks (ms)=5773
Total vcore-milliseconds taken by all map tasks=5326
Total vcore-milliseconds taken by all reduce tasks=5773
Total megabyte-milliseconds taken by all map tasks=5453824
Total megabyte-milliseconds taken by all reduce tasks=5911552
Map-Reduce Framework
  Map input records=8
  Map output records=72
  Map output bytes=1056
  Map output materialized bytes=56
  Input split bytes=110
  Combine input records=72
  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)=140
  CPU time spent (ms)=1490
  Physical memory (bytes) snapshot=349900800
  Virtual memory (bytes) snapshot=4127399936
  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=48
File Output Format Counters
  Bytes Written=29
The standard deviation is: 0.0
```

Output

acadgild@localhost:~

```
[acadgild@localhost ~]$ hdfs dfs -cat WordStandardDeviation/part-r-00000
18/01/28 16:13:43 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using
count 24
length 24
square 24
[acadgild@localhost ~]$
```