

4. Problem Statement

Task 1:

Execute **WordMedian** , **WordMean** , **WordStandardDeviation** programs using **hadoop-mapreduce-examples-2.9.0.jar** file present in your AcadGild VM.

Refer path below.

/home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce

WordMedian

Uploaded words.txt to hdfs /user/acadgild/words.

```
[acadgild@localhost mapreduce]$ ls
hadoop-mapreduce-client-app-2.6.5.jar      hadoop-mapreduce-client-hs-2.6.5.jar      hadoop-mapreduce-client-jobclient-2.6.5-tests.jar  lib      tony.txt
hadoop-mapreduce-client-common-2.6.5.jar   hadoop-mapreduce-client-hs-plugins-2.6.5.jar  hadoop-mapreduce-client-shuffle-2.6.5.jar          lib-examples  wordmedianout.txt
hadoop-mapreduce-client-core-2.6.5.jar     hadoop-mapreduce-client-jobclient-2.6.5.jar  hadoop-mapreduce-examples-2.6.5.jar               sources      words.txt
[acadgild@localhost mapreduce]$ hadoop fs -put words.txt /user/acadgild/words.txt
18/09/21 02:40:32 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
```

Then Executed command to median words from the list

```
[acadgild@localhost mapreduce]$ ls
hadoop-mapreduce-client-app-2.6.5.jar      hadoop-mapreduce-client-hs-2.6.5.jar      hadoop-mapreduce-client-jobclient-2.6.5-tests.jar  lib      tony.txt
hadoop-mapreduce-client-common-2.6.5.jar   hadoop-mapreduce-client-hs-plugins-2.6.5.jar  hadoop-mapreduce-client-shuffle-2.6.5.jar          lib-examples  wordmedianout.txt
hadoop-mapreduce-client-core-2.6.5.jar     hadoop-mapreduce-client-jobclient-2.6.5.jar  hadoop-mapreduce-examples-2.6.5.jar               sources      words.txt
[acadgild@localhost mapreduce]$ hadoop fs -put words.txt /user/acadgild/words.txt
18/09/21 02:40:32 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
[acadgild@localhost mapreduce]$ hadoop jar hadoop-mapreduce-examples-2.6.5.jar wordmedian words.txt wordmedianout.txt
18/09/21 02:41:21 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/09/21 02:41:23 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy th
18/09/21 02:41:23 INFO input.FileInputFormat: Total input paths to process : 1
18/09/21 02:41:23 INFO mapreduce.JobSubmitter: number of splits:1
18/09/21 02:41:24 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1537476369822_0003
18/09/21 02:41:24 INFO impl.YarnClientImpl: Submitted application application_1537476369822_0003
18/09/21 02:41:24 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1537476369822_0003/
18/09/21 02:41:24 INFO mapreduce.Job: Running job: job_1537476369822_0003
18/09/21 02:41:38 INFO mapreduce.Job: Job job_1537476369822_0003 running in uber mode : false
18/09/21 02:41:38 INFO mapreduce.Job: map 0% reduce 0%
18/09/21 02:41:47 INFO mapreduce.Job: map 100% reduce 0%
18/09/21 02:41:56 INFO mapreduce.Job: map 100% reduce 100%
18/09/21 02:41:57 INFO mapreduce.Job: Job job_1537476369822_0003 completed successfully
18/09/21 02:41:57 INFO mapreduce.Job: Counters: 49
  File System Counters
    FILE: Number of bytes read=316
    FILE: Number of bytes written=215749
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=4863076
    HDFS: Number of bytes written=220
    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)=6789
    Total time spent by all reduces in occupied slots (ms)=6669
    Total time spent by all map tasks (ms)=6789
    Total time spent by all reduce tasks (ms)=6669
    Total vcore-milliseconds taken by all map tasks=6789
    Total vcore-milliseconds taken by all reduce tasks=6669
    Total megabyte-milliseconds taken by all map tasks=6951936
    Total megabyte-milliseconds taken by all reduce tasks=6829056
  Map-Reduce Framework
    Map input records=466544
    Map output records=466544
    Map output bytes=3732352
    Map output materialized bytes=316
    Input split bytes=110
    Combine input records=466544
    Combine output records=31
    Reduce input groups=31
    Reduce shuffle bytes=316
    Reduce input records=31
    Reduce output records=31
    Spilled Records=62
    Shuffled Maps =1
    Failed Shuffles=1
    Merged Map outputs=1
    GC time elapsed (ms)=162
    CPU time spent (ms)=3060
    Physical memory (bytes) snapshot=300437504
    Virtual memory (bytes) snapshot=418216704
    Total committed heap usage (bytes)=170004480
  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=4862966
  File Output Format Counters
    Bytes Written=220
The median is: 9
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost mapreduce]$
```

Results:

```
[acadgild@localhost mapreduce]$ hadoop fs -cat /user/acadgild/wordmedianout.txt/part-r-00000
18/09/21 02:46:44 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
1 27
2 635
3 4710
4 11169
5 22950
6 39518
7 52092
8 61018
9 61755
10 54321
11 46410
12 37525
13 27924
14 19253
15 12148
16 7115
17 3982
18 2003
19 1053
20 506
21 238
22 102
23 49
24 18
25 7
26 2
27 2
28 2
29 2
31 1
45 1
[acadgild@localhost mapreduce]$ |
```

WordMean

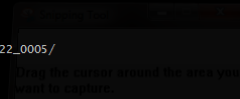
```
[acadgild@localhost mapreduce]$ hadoop jar hadoop-mapreduce-examples-2.6.0.jar wordmean words.txt wordmeanout.txt
18/09/21 02:51:12 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/09/21 02:51:13 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/09/21 02:51:15 INFO input.FileInputFormat: Total input paths to process : 1
18/09/21 02:51:15 INFO mapreduce.JobSubmitter: number of splits:1
18/09/21 02:51:15 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1537476369822_0004
18/09/21 02:51:16 INFO impl.YarnClientImpl: Submitted application application_1537476369822_0004
18/09/21 02:51:16 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1537476369822_0004/
18/09/21 02:51:16 INFO mapreduce.Job: Running job: job_1537476369822_0004
18/09/21 02:51:27 INFO mapreduce.Job: Job job_1537476369822_0004 running in uber mode : false
18/09/21 02:51:27 INFO mapreduce.Job: map 0% reduce 0%
18/09/21 02:51:36 INFO mapreduce.Job: map 100% reduce 0%
18/09/21 02:51:44 INFO mapreduce.Job: map 100% reduce 100%
18/09/21 02:51:45 INFO mapreduce.Job: Job job_1537476369822_0004 completed successfully
18/09/21 02:51:46 INFO mapreduce.Job: Counters: 49
  File System Counters
    FILE: Number of bytes read=39
    FILE: Number of bytes written=215455
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=4863076
    HDFS: Number of bytes written=28
    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)=7347
    Total time spent by all reduces in occupied slots (ms)=5624
    Total time spent by all map tasks (ms)=7347
    Total time spent by all reduce tasks (ms)=5624
    Total vcore-milliseconds taken by all map tasks=7347
    Total vcore-milliseconds taken by all reduce tasks=5624
    Total megabyte-milliseconds taken by all map tasks=7523328
    Total megabyte-milliseconds taken by all reduce tasks=5758976
  Map-Reduce Framework
    Map input records=466544
    Map output records=933088
    Map output bytes=1352976
    Map output materialized bytes=39
    Input split bytes=110
    Combine input records=933088
    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)=164
    CPU time spent (ms)=3550
    Physical memory (bytes) snapshot=302465024
    Virtual memory (bytes) snapshot=4118224896
    Total committed heap usage (bytes)=170004480
  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=4862966
  File Output Format Counters
    Bytes Written=28
The mean is: 9.423381286795913
[acadgild@localhost mapreduce]$ |
```

Results

```
[acadgild@localhost mapreduce]$ hadoop fs -cat /user/acadgild/wordmeanout.txt/part-r-00000
18/09/21 02:53:24 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
count 466544
length 4396422
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost mapreduce]$ |
```

WordStandardDeviation

```
[acadgild@localhost mapreduce]$ hadoop jar hadoop-mapreduce-examples-2.6.5.jar wordstandarddeviation words.txt wordsdout.txt
18/09/21 02:54:57 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/09/21 02:54:58 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/09/21 02:55:00 INFO input.FileInputFormat: Total input paths to process : 1
18/09/21 02:55:00 INFO mapreduce.JobSubmitter: number of splits=1
18/09/21 02:55:00 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1537476369822_0005
18/09/21 02:55:01 INFO impl.YarnClientImpl: Submitted application application_1537476369822_0005
18/09/21 02:55:01 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1537476369822_0005/
18/09/21 02:55:01 INFO mapreduce.Job: Running job: job_1537476369822_0005
18/09/21 02:55:11 INFO mapreduce.Job: Job job_1537476369822_0005 running in uber mode : false
18/09/21 02:55:11 INFO mapreduce.Job: map 0% reduce 0%
18/09/21 02:55:21 INFO mapreduce.Job: map 100% reduce 0%
18/09/21 02:55:29 INFO mapreduce.Job: map 100% reduce 100%
18/09/21 02:55:29 INFO mapreduce.Job: Job job_1537476369822_0005 completed successfully
18/09/21 02:55:30 INFO mapreduce.Job: Counters: 49
  File System Counters
    FILE: Number of bytes read=56
    FILE: Number of bytes written=215645
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=4863076
    HDFS: Number of bytes written=44
    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)=7513
    Total time spent by all reduces in occupied slots (ms)=5730
    Total time spent by all map tasks (ms)=7513
    Total time spent by all reduce tasks (ms)=5730
    Total vcore-milliseconds taken by all map tasks=7513
    Total vcore-milliseconds taken by all reduce tasks=5730
    Total megabyte-milliseconds taken by all map tasks=7693312
    Total megabyte-milliseconds taken by all reduce tasks=5867320
  Map-Reduce Framework
    Map input records=466544
    Map output records=1399632
    Map output bytes=20527936
    Map output materialized bytes=56
    Input split bytes=110
    Combine input records=1399632
    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)=178
    CPU time spent (ms)=3840
    Physical memory (bytes) snapshot=301592576
    Virtual memory (bytes) snapshot=4118220800
    Total committed heap usage (bytes)=170004480
  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=4862966
  File Output Format Counters
    Bytes Written=44
The standard deviation is: 3.049131835934211
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost mapreduce]$
```



Results

```
[acadgild@localhost mapreduce]$ hadoop fs -cat /user/acadgild/wordsdout.txt/part-r-00000
count  466544
length 4396422
square 45766716
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost mapreduce]$
```