

Task 1:

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

Solution :

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

```
18/06/26 11:36:08 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/06/26 11:36:10 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/06/26 11:36:20 INFO input.FileInputFormat: Total input paths to process : 1
18/06/26 11:36:22 INFO mapreduce.JobSubmitter: number of splits:1
18/06/26 11:36:24 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1529992151660_0001
18/06/26 11:36:27 INFO impl.YarnClientImpl: Submitted application application_1529992151660_0001
18/06/26 11:36:28 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1529992151660_0001/
18/06/26 11:36:28 INFO mapreduce.Job: Running job: job_1529992151660_0001
18/06/26 11:36:56 INFO mapreduce.Job: Job job_1529992151660_0001 running in uber mode : false
18/06/26 11:36:56 INFO mapreduce.Job: map 0% reduce 0%
18/06/26 11:37:09 INFO mapreduce.Job: map 100% reduce 0%
18/06/26 11:37:26 INFO mapreduce.Job: map 100% reduce 100%
18/06/26 11:37:27 INFO mapreduce.Job: Job job_1529992151660_0001 completed successfully
18/06/26 11:37:27 INFO mapreduce.Job: Counters: 49
  File System Counters
    FILE: Number of bytes read=39
    FILE: Number of bytes written=215465
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=144
    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)=8585
    Total time spent by all reduces in occupied slots (ms)=13651
    Total time spent by all map tasks (ms)=8585
    Total time spent by all reduce tasks (ms)=13651
```

```
    Total time spent by all reduces in occupied slots (ms)=13651
    Total time spent by all map tasks (ms)=8585
    Total time spent by all reduce tasks (ms)=13651
    Total vcore-milliseconds taken by all map tasks=8585
    Total vcore-milliseconds taken by all reduce tasks=13651
    Total megabyte-milliseconds taken by all map tasks=8791040
    Total megabyte-milliseconds taken by all reduce tasks=13978624
  Map-Reduce Framework
    Map input records=1
    Map output records=10
    Map output bytes=145
    Map output materialized bytes=39
    Input split bytes=115
    Combine input records=10
    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)=283
    CPU time spent (ms)=2320
    Physical memory (bytes) snapshot=301809664
    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=29
  File Output Format Counters
```

```

        Merged Map outputs=1
        GC time elapsed (ms)=283
        CPU time spent (ms)=2320
        Physical memory (bytes) snapshot=301809664
        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=29
File Output Format Counters
        Bytes Written=18
The mean is: 4.8
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$ hdfs dfs -cat /hadoopdata/wordmean/myoutput/*

```

```

cat: /hadoopdata/wordmean/myoutput/*: No such file or directory
[acadgild@localhost ~]$ hdfs dfs -cat /hadoopdata/wordmean/myoutput/*
18/06/26 11:46:31 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
count 5
length 24
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$

```

WORDEMEDIAN:

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

```

found 2 items
drwxr-xr-x - acadgild supergroup 0 2018-06-25 12:25 /hadoopdata/wordcount
drwxr-xr-x - acadgild supergroup 0 2018-06-26 11:49 /hadoopdata/wordmean
[acadgild@localhost ~]$ hadoop jar $HADOOP_HOME/share/hadoop/mapreduce/hadoop-mapreduce-examples-2.6.5.jar wordmedian /hadoopdata/wordmean/test.txt /hadoopdata/wordmean/myoutput
18/06/26 11:50:13 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/06/26 11:50:14 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/06/26 11:50:15 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
18/06/26 11:50:17 INFO input.FileInputFormat: Total input paths to process : 1
18/06/26 11:50:17 INFO mapreduce.JobSubmitter: number of splits:1
18/06/26 11:50:17 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1529992151660_0002
18/06/26 11:50:17 INFO impl.YarnClientImpl: Submitted application application_1529992151660_0002
18/06/26 11:50:18 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1529992151660_0002/
18/06/26 11:50:18 INFO mapreduce.Job: Running job: job_1529992151660_0002
18/06/26 11:50:32 INFO mapreduce.Job: Job job_1529992151660_0002 running in uber mode : false
18/06/26 11:50:32 INFO mapreduce.Job: map 0% reduce 0%
18/06/26 11:50:42 INFO mapreduce.Job: map 100% reduce 0%
18/06/26 11:50:52 INFO mapreduce.Job: map 100% reduce 100%
18/06/26 11:50:53 INFO mapreduce.Job: Job job_1529992151660_0002 completed successfully
18/06/26 11:50:54 INFO mapreduce.Job: Counters: 49
File System Counters
        FILE: Number of bytes read=46
        FILE: Number of bytes written=215215
        FILE: Number of read operations=0
        FILE: Number of large read operations=0
        FILE: Number of write operations=0
        HDFS: Number of bytes read=144
        HDFS: Number of bytes written=17
        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)=7969

```

```

    Launched map tasks=1
    Launched reduce tasks=1
    Data-local map tasks=1
    Total time spent by all maps in occupied slots (ms)=7969
    Total time spent by all reduces in occupied slots (ms)=8440
    Total time spent by all map tasks (ms)=7969
    Total time spent by all reduce tasks (ms)=8440
    Total vcore-milliseconds taken by all map tasks=7969
    Total vcore-milliseconds taken by all reduce tasks=8440
    Total megabyte-milliseconds taken by all map tasks=8160256
    Total megabyte-milliseconds taken by all reduce tasks=8642560
Map-Reduce Framework
    Map input records=1
    Map output records=5
    Map output bytes=40
    Map output materialized bytes=46
    Input split bytes=115
    Combine input records=5
    Combine output records=4
    Reduce input groups=4
    Reduce shuffle bytes=46
    Reduce input records=4
    Reduce output records=4
    Spilled Records=8
    Shuffled Maps =1
    Failed Shuffles=0
    Merged Map outputs=1
    GC time elapsed (ms)=209
    CPU time spent (ms)=2300
    Physical memory (bytes) snapshot=297922560
    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

```

```

    Input split bytes=115
    Combine input records=5
    Combine output records=4
    Reduce input groups=4
    Reduce shuffle bytes=46
    Reduce input records=4
    Reduce output records=4
    Spilled Records=8
    Shuffled Maps =1
    Failed Shuffles=0
    Merged Map outputs=1
    GC time elapsed (ms)=209
    CPU time spent (ms)=2300
    Physical memory (bytes) snapshot=297922560
    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=29
File Output Format Counters
    Bytes Written=17
The median is: 4
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$ hdfs dfs -cat /hadoopdata/wordmean/myoutpput/*
18/06/26 11:51:08 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
1      1
4      2
5      1
10     1
[acadgild@localhost ~]$ █

```

WORDSTANDARDDEVIATION:

A map/reduce program that counts the standarddeviation of the length of the words in the input files.

```

You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$ hadoop jar $HADOOP_HOME/share/hadoop/mapreduce/hadoop-mapreduce-examples-2.6.5.jar wordstandarddeviation /hadoop
ata/wordmean/test.txt /hadoopdata/wordmean/myoutpput
18/06/26 11:53:50 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where
applicable
18/06/26 11:53:52 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/06/26 11:53:54 INFO input.FileInputFormat: Total input paths to process : 1
18/06/26 11:53:54 INFO mapreduce.JobSubmitter: number of splits:1
18/06/26 11:53:54 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1529992151660_0003
18/06/26 11:53:55 INFO impl.YarnClientImpl: Submitted application application_1529992151660_0003
18/06/26 11:53:55 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1529992151660_0003/
18/06/26 11:53:55 INFO mapreduce.Job: Running job: job_1529992151660_0003
18/06/26 11:54:08 INFO mapreduce.Job: Job job_1529992151660_0003 running in uber mode : false
18/06/26 11:54:08 INFO mapreduce.Job: map 0% reduce 0%
18/06/26 11:54:17 INFO mapreduce.Job: map 100% reduce 0%
18/06/26 11:54:29 INFO mapreduce.Job: map 100% reduce 100%
18/06/26 11:54:29 INFO mapreduce.Job: Job job_1529992151660_0003 completed successfully
18/06/26 11:54:29 INFO mapreduce.Job: Counters: 49
File System Counters
  FILE: Number of bytes read=56
  FILE: Number of bytes written=215659
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=144
  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)=6957
  Total time spent by all reduces in occupied slots (ms)=7662
  Total time spent by all map tasks (ms)=6957
  Total time spent by all reduce tasks (ms)=7662
  Total vcore-milliseconds taken by all map tasks=6957

```

```

  Total time spent by all reduce tasks (ms)=7662
  Total vcore-milliseconds taken by all map tasks=6957
  Total vcore-milliseconds taken by all reduce tasks=7662
  Total megabyte-milliseconds taken by all map tasks=7123968
  Total megabyte-milliseconds taken by all reduce tasks=7845888
Map-Reduce Framework
  Map input records=1
  Map output records=15
  Map output bytes=220
  Map output materialized bytes=56
  Input split bytes=115
  Combine input records=15
  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)=187
  CPU time spent (ms)=2090
  Physical memory (bytes) snapshot=296824832
  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=29
File Output Format Counters
  Bytes Written=29
The standard deviation is: 2.925747767665559

```

```

  CONNECTION=0
  IO_ERROR=0
  WRONG_LENGTH=0
  WRONG_MAP=0
  WRONG_REDUCE=0
File Input Format Counters
  Bytes Read=29
File Output Format Counters
  Bytes Written=29
The standard deviation is: 2.925747767665559
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$ hdfs dfs -cat /hadoopdata/wordmean/myoutpput/*
18/06/26 11:54:56 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where
applicable
count      5
length    24
square    158
[acadgild@localhost ~]$ █

```