

## Day 3

### Task 1:

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

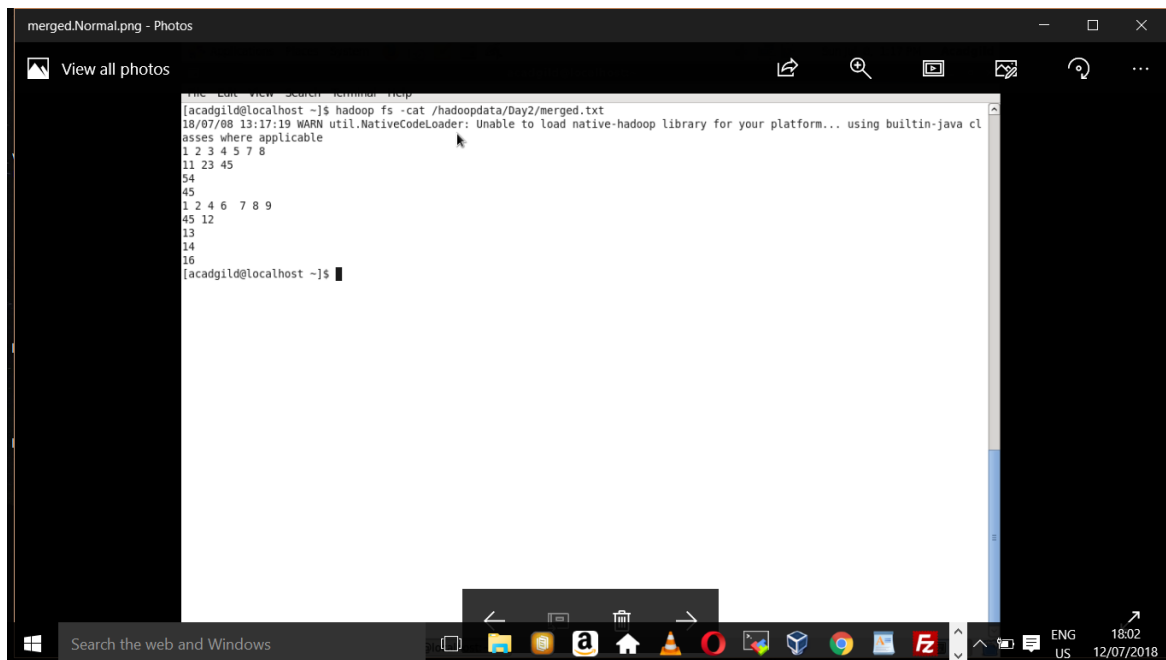
Ans:

Input file contents before running each of above commands:

```
hadoop fs -cat /hadoopdata/Day2/merged.txt
```

(Displays the complete content of the file passed onto terminal)

ScreenShot(s):



## 1: WordMedian

```
hadoop jar hadoop-mapreduce-examples-2.6.5.jar wordmedian  
/hadoopdata/Day2/merged.txt /myoutputmedian
```

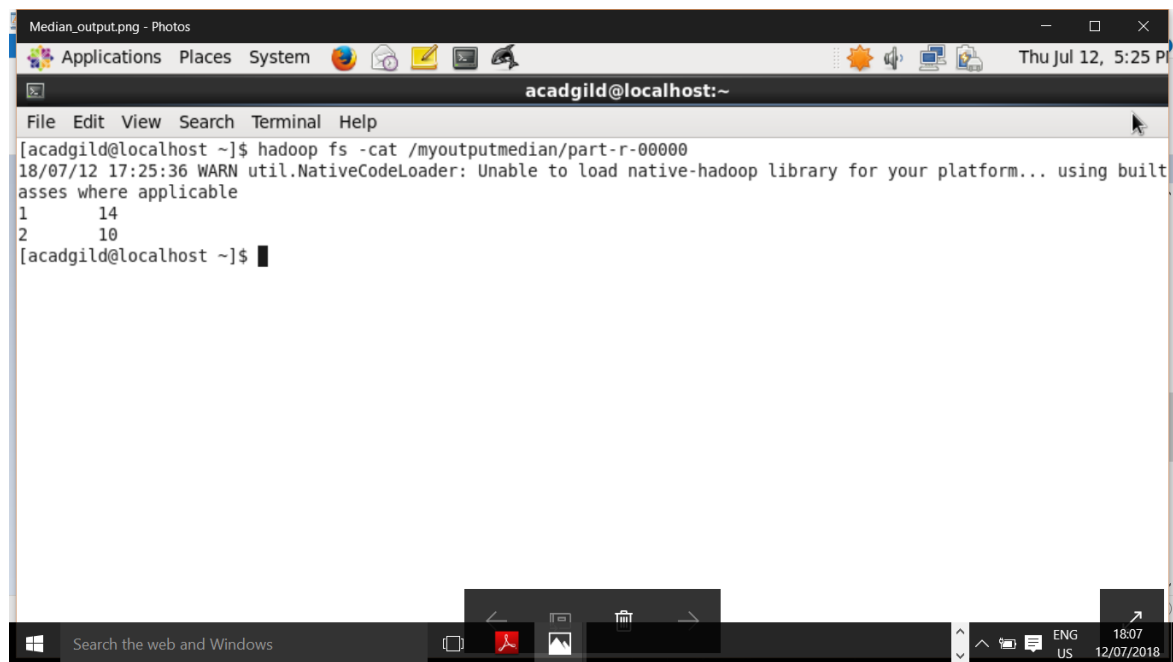
(A map/reduce program that counts the median length of the words in the input file(s).)

### Output:

```
Combine output records=2  
Reduce input groups=2  
Reduce shuffle bytes=26  
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)=523  
CPU time spent (ms)=7160  
Physical memory (bytes) snapshot=407154688  
Virtual memory (bytes) snapshot=4167270400  
Total committed heap usage (bytes)=280494080  
  
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=70  
File Output Format Counters  
Bytes Written=10  
  
The median is: 1  
[acadgild@localhost ~]$
```



### Out-File:

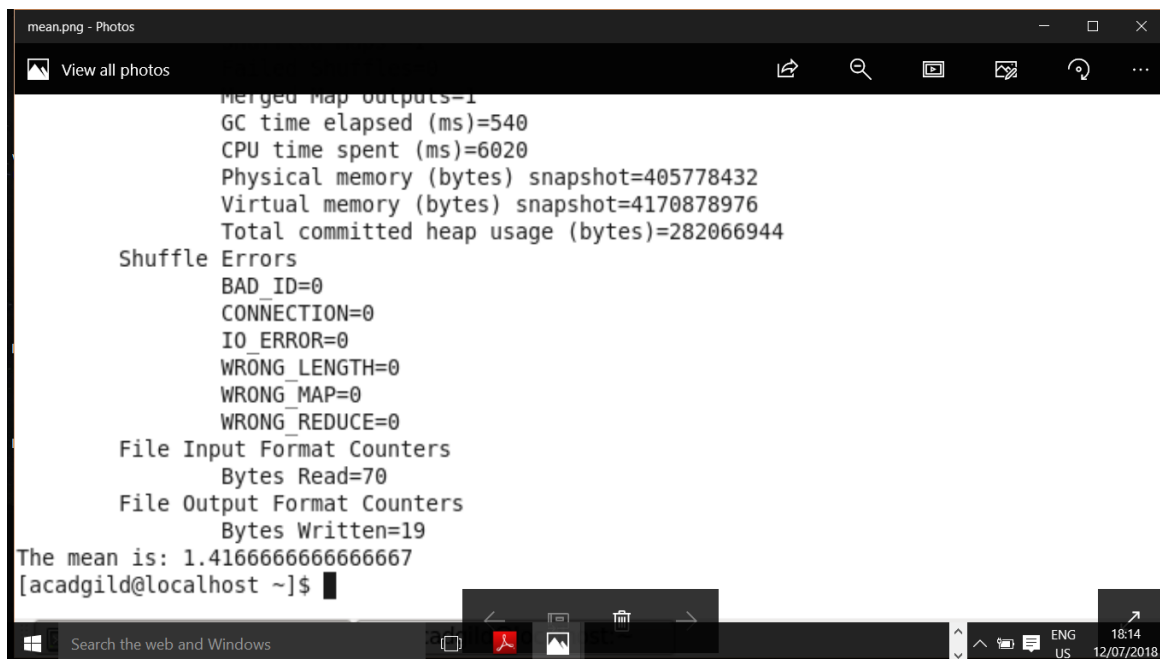
A screenshot of a Windows 10 desktop. At the top, a taskbar shows the Start button, a search bar with the text "Search the web and Windows", and several pinned application icons including File Explorer, Microsoft Edge, and a terminal. The terminal window is open, displaying the command prompt for a user named 'acadgild' on a 'localhost'. The command executed is 'hadoop fs -cat /myoutputmedian/part-r-00000'. The output shows a warning message from 'util.NativeCodeLoader' and two lines of data: '1 14' and '2 10'. The Windows taskbar at the bottom shows the system tray with the date '12/07/2018' and time '18:07'.

## 2: WordMean

```
hadoop jar hadoop-mapreduce-examples-2.6.5.jar wordmean  
/hadoopdata/Day2/merged.txt /myoutput_mean
```

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

ScreenShot(s):



```
mean.png - Photos
View all photos

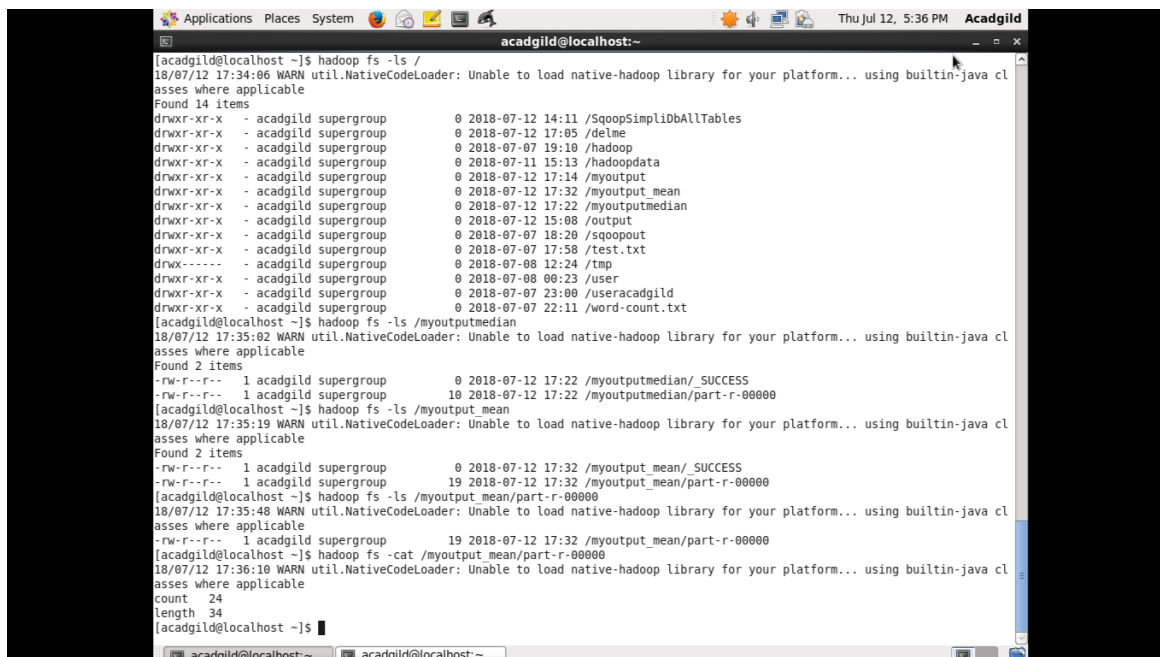
Merged Map outputs=1
GC time elapsed (ms)=540
CPU time spent (ms)=6020
Physical memory (bytes) snapshot=405778432
Virtual memory (bytes) snapshot=4170878976
Total committed heap usage (bytes)=282066944

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=70
File Output Format Counters
Bytes Written=19

The mean is: 1.4166666666666667
[acadgild@localhost ~]$
```

out-file:



```
Applications Places System Thu Jul 12, 5:36 PM acadgild
acadgild@localhost:~
[acadgild@localhost ~]$ hadoop fs -ls /
18/07/12 17:34:06 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
Found 14 items
drwxr-xr-x - acadgild supergroup 0 2018-07-12 14:11 /SqoopSimlibAllTables
drwxr-xr-x - acadgild supergroup 0 2018-07-12 17:05 /delme
drwxr-xr-x - acadgild supergroup 0 2018-07-07 19:10 /hadoop
drwxr-xr-x - acadgild supergroup 0 2018-07-11 15:13 /hadoopdata
drwxr-xr-x - acadgild supergroup 0 2018-07-12 17:14 /myoutput
drwxr-xr-x - acadgild supergroup 0 2018-07-12 17:32 /myoutput_mean
drwxr-xr-x - acadgild supergroup 0 2018-07-12 17:22 /myoutputmedian
drwxr-xr-x - acadgild supergroup 0 2018-07-12 15:08 /output
drwxr-xr-x - acadgild supergroup 0 2018-07-07 18:20 /sqoopout
drwxr-xr-x - acadgild supergroup 0 2018-07-07 17:50 /test.txt
drwxr-xr-x - acadgild supergroup 0 2018-07-08 12:24 /tmp
drwxr-xr-x - acadgild supergroup 0 2018-07-08 00:23 /user
drwxr-xr-x - acadgild supergroup 0 2018-07-07 23:00 /useracadgild
drwxr-xr-x - acadgild supergroup 0 2018-07-07 22:11 /word-count.txt
[acadgild@localhost ~]$ hadoop fs -ls /myoutputmedian
18/07/12 17:35:02 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
Found 2 items
-rw-r--r-- 1 acadgild supergroup 0 2018-07-12 17:22 /myoutputmedian/ SUCCESS
-rw-r--r-- 1 acadgild supergroup 10 2018-07-12 17:22 /myoutputmedian/part-r-000000
[acadgild@localhost ~]$ hadoop fs -ls /myoutput_mean
18/07/12 17:35:19 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
Found 2 items
-rw-r--r-- 1 acadgild supergroup 0 2018-07-12 17:32 /myoutput_mean/ SUCCESS
-rw-r--r-- 1 acadgild supergroup 19 2018-07-12 17:32 /myoutput_mean/part-r-000000
[acadgild@localhost ~]$ hadoop fs -ls /myoutput_mean/part-r-000000
18/07/12 17:35:40 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
-rw-r--r-- 1 acadgild supergroup 19 2018-07-12 17:32 /myoutput_mean/part-r-000000
[acadgild@localhost ~]$ hadoop fs -cat /myoutput_mean/part-r-000000
18/07/12 17:36:10 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
count 24
length 34
[acadgild@localhost ~]$
```

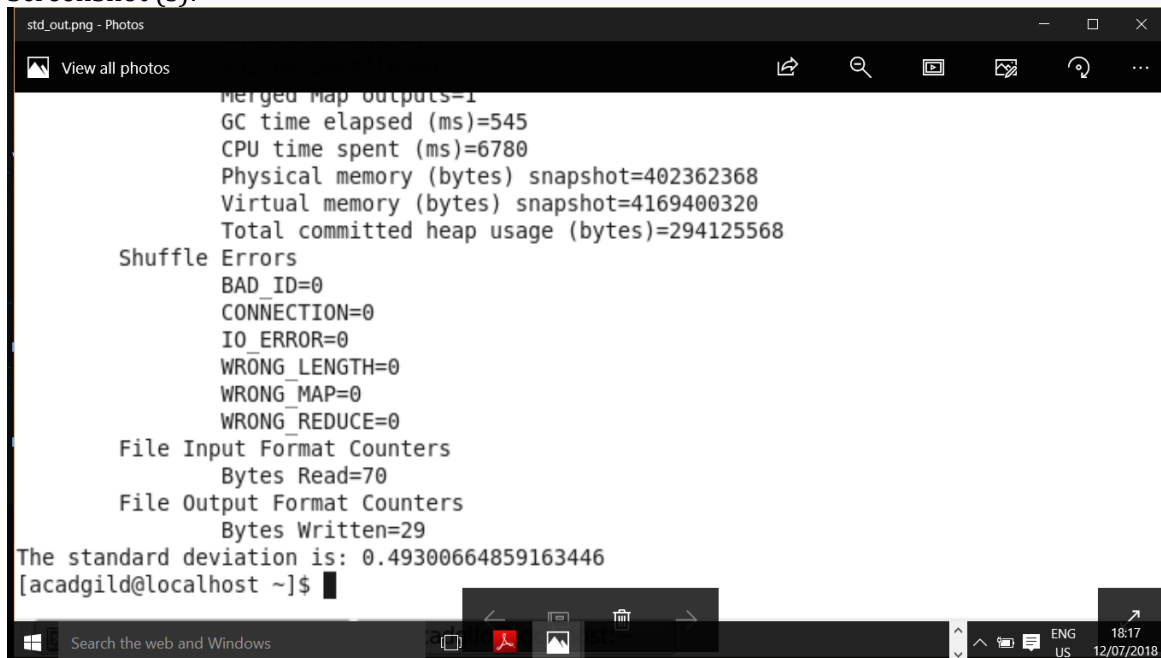
### 3:WordStandardDeviation

hadoop jar hadoop-mapreduce-examples-2.6.5.jar wordstandarddeviation  
/hadoopdata/Day2/merged.txt /myoutput\_std

(A map/reduce program that counts the standarddeviation of the length of the words in the

input file(s) )

ScreenShot (s):



A screenshot of a Windows Photos application window titled 'std\_out.png - Photos'. The window displays the output of a Hadoop job. The output text is as follows:

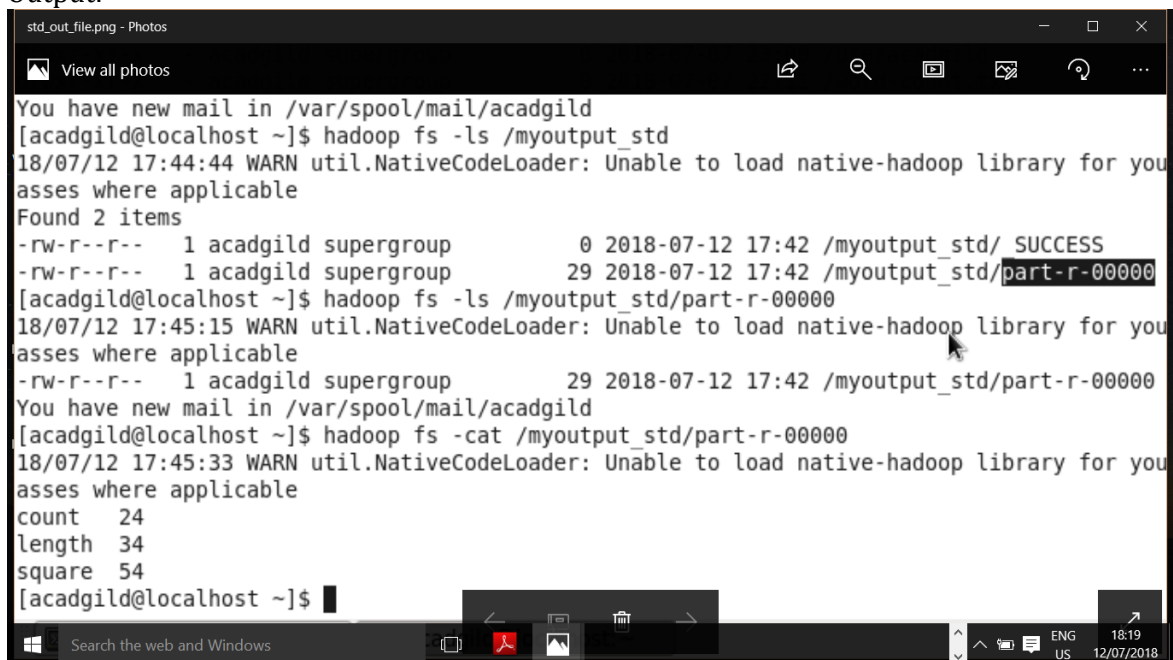
```
Merged Map outputs=1
GC time elapsed (ms)=545
CPU time spent (ms)=6780
Physical memory (bytes) snapshot=402362368
Virtual memory (bytes) snapshot=4169400320
Total committed heap usage (bytes)=294125568

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=70
File Output Format Counters
  Bytes Written=29
The standard deviation is: 0.49300664859163446
[acadgild@localhost ~]$
```

The terminal window is part of a Windows desktop environment, with the taskbar at the bottom showing the Start button, search bar, and several open applications. The system clock in the bottom right corner indicates the date as 12/07/2018 and the time as 18:17.

Output:



A screenshot of a Windows Photos application window titled 'std\_out\_file.png - Photos'. The window displays the output of Hadoop file system commands. The output text is as follows:

```
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$ hadoop fs -ls /myoutput_std
18/07/12 17:44:44 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your
asses where applicable
Found 2 items
-rw-r--r-- 1 acadgild supergroup          0 2018-07-12 17:42 /myoutput_std/ SUCCESS
-rw-r--r-- 1 acadgild supergroup        29 2018-07-12 17:42 /myoutput_std/part-r-00000
[acadgild@localhost ~]$ hadoop fs -ls /myoutput_std/part-r-00000
18/07/12 17:45:15 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your
asses where applicable
-rw-r--r-- 1 acadgild supergroup        29 2018-07-12 17:42 /myoutput_std/part-r-00000
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$ hadoop fs -cat /myoutput_std/part-r-00000
18/07/12 17:45:33 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your
asses where applicable
count  24
length 34
square 54
[acadgild@localhost ~]$
```

The terminal window is part of a Windows desktop environment, with the taskbar at the bottom showing the Start button, search bar, and several open applications. The system clock in the bottom right corner indicates the date as 12/07/2018 and the time as 18:19.