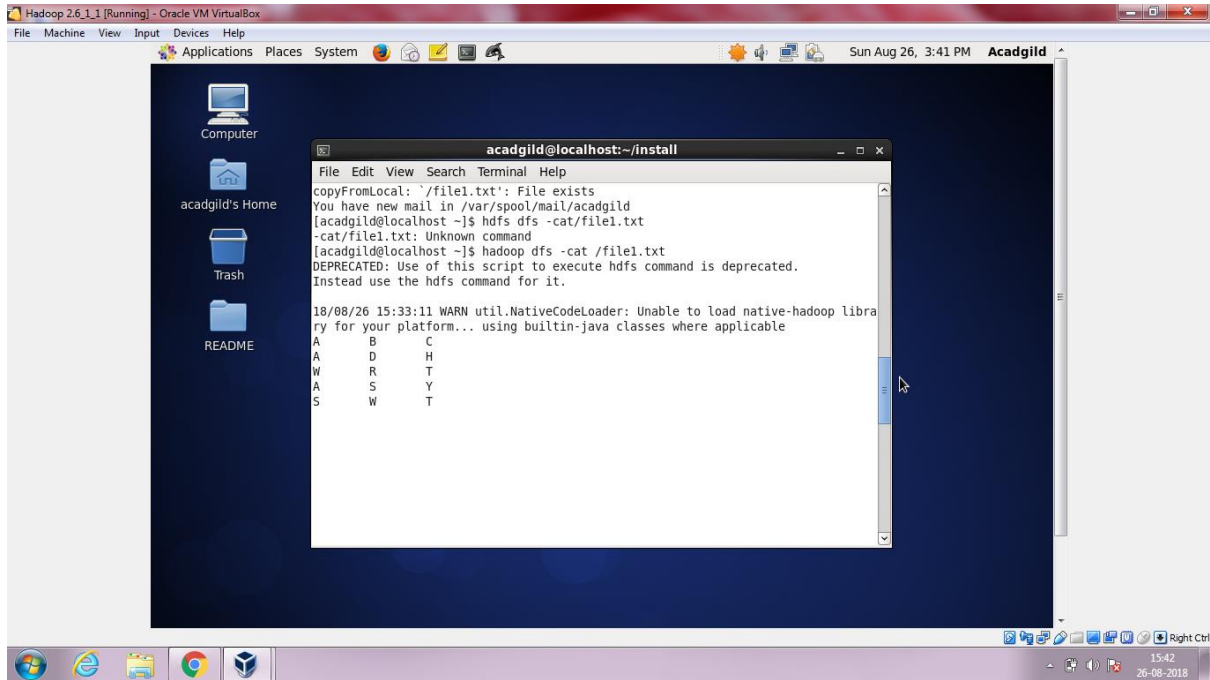


ASSIGNMENT 3

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

Input file :file1.txt



The screenshot shows a terminal window titled 'acadgild@localhost:~/install' within an Oracle VM VirtualBox environment. The terminal displays the following commands and output:

```
copyFromLocal: '/file1.txt': File exists
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$ hdfs dfs -cat /file1.txt
-cat /file1.txt: Unknown command
[acadgild@localhost ~]$ hadoop dfs -cat /file1.txt
DEPRECATED: Use of this script to execute hdfs command is deprecated.
Instead use the hdfs command for it.

18/08/26 15:33:11 WARN util.NativeCodeLoader: Unable to load native-hadoop library
for your platform... using builtin-java classes where applicable
A      B      C
A      D      H
W      R      T
A      S      Y
S      W      T
```

The output shows a 5x3 grid of characters. The first command 'hdfs dfs -cat /file1.txt' failed with 'Unknown command'. The second command 'hadoop dfs -cat /file1.txt' succeeded and displayed the file contents. A warning message about the native-hadoop library is also visible.

Calculate wordmedian:

```
Hadoop 2.6.1.1 [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
acacgild@localhost:~/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce
File Edit View Search Terminal Help
[acacgild@localhost mapreduce]$ hadoop jar hadoop-mapreduce-examples-2.6.5.jar wordmedian /file1.txt wordmedian
18/08/26 15:55:44 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/08/26 15:55:46 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/08/26 15:55:49 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
18/08/26 15:55:51 INFO input.FileInputFormat: Total input paths to process : 1
18/08/26 15:55:52 INFO mapreduce.JobSubmitter: number of splits:1
18/08/26 15:55:53 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1535277690582_0001
18/08/26 15:55:54 INFO impl.YarnClientImpl: Submitted application application_1535277690582_0001
18/08/26 15:55:55 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1535277690582_0001/
18/08/26 15:55:55 INFO mapreduce.Job: Running job: job_1535277690582_0001
18/08/26 15:56:24 INFO mapreduce.Job: Job job_1535277690582_0001 running in uber mode : false
18/08/26 15:56:24 INFO mapreduce.Job: map 0% reduce 0%
18/08/26 15:56:40 INFO mapreduce.Job: map 100% reduce 0%
18/08/26 15:56:56 INFO mapreduce.Job: map 100% reduce 100%
18/08/26 15:56:57 INFO mapreduce.Job: Job job_1535277690582_0001 completed successfully
18/08/26 15:56:58 INFO mapreduce.Job: Counters: 49
    File System Counters
      FILE: Number of bytes read=16
      FILE: Number of bytes written=215107
      FILE: Number of read operations=0
      FILE: Number of large read operations=0
      FILE: Number of write operations=0
      HDFS: Number of bytes read=147
      HDFS: Number of bytes written=5
      HDFS: Number of read operations=6
      HDFS: Number of large read operations=0
```

```
Hadoop 2.6.1.1 [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Total megabyte-milliseconds taken by all map tasks=13274112
Total megabyte-milliseconds taken by all reduce tasks=13936640
Map-Reduce Framework
  Map input records=26
  Map output records=15
  Map output bytes=120
  Map output materialized bytes=16
  Input split bytes=96
  Combine input records=15
  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)=371
  CPU time spent (ms)=3500
  Physical memory (bytes) snapshot=296259584
  Virtual memory (bytes) snapshot=4118220800
  Total committed heap usage (bytes)=170064480
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=51
File Output Format Counters
  Bytes Written=5
The median is: 1
You have new mail in /var/spool/mail/acacgild
[acacgild@localhost mapreduce]$
acacgild@localhost:~/...
15:59
26-08-2018
```

```
Hadoop 2.6.1.1 [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help

Map output records=15
Map output bytes=120
Map output materialized bytes=16
Input split bytes=96
Combine input records=15
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)=371
CPU time spent (ms)=3500
Physical memory (bytes) snapshot=296259584
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=51
File Output Format Counters
Bytes Written=5

The median is: 1
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost mapreduce]$ hdfs dfs -cat wordmedian/part-r-00000
18/08/26 16:00:15 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
1
15
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost mapreduce]$
```

Calculate wordmean:

```
Hadoop 2.6.1.1 [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help

[acadgild@localhost mapreduce]$ hadoop jar hadoop-mapreduce-examples-2.6.5.jar wordmean /file1.txt wordmean
18/08/26 16:01:55 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
18/08/26 16:01:58 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/08/26 16:02:01 INFO input.FileInputFormat: Total input paths to process : 1
18/08/26 16:02:01 INFO mapreduce.JobSubmitter: number of splits:1
18/08/26 16:02:02 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1535277690582_0002
18/08/26 16:02:02 INFO impl.YarnClientImpl: Submitted application application_1535277690582_0002
18/08/26 16:02:02 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1535277690582_0002/
18/08/26 16:02:02 INFO mapreduce.Job: Running job: job_1535277690582_0002
18/08/26 16:02:20 INFO mapreduce.Job: Job job_1535277690582_0002 running in uber mode : false
18/08/26 16:02:20 INFO mapreduce.Job: map 0% reduce 0%
18/08/26 16:02:34 INFO mapreduce.Job: map 100% reduce 0%
18/08/26 16:02:47 INFO mapreduce.Job: map 100% reduce 100%
18/08/26 16:02:49 INFO mapreduce.Job: Job job_1535277690582_0002 completed successfully
18/08/26 16:02:49 INFO mapreduce.Job: Counters: 49

File System Counters
FILE: Number of bytes read=39
FILE: Number of bytes written=215413
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=147
HDFS: Number of bytes written=19
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)=10915
Total time spent by all reduces in occupied slots (ms)=10716
Total time spent by all map tasks (ms)=10915
Total time spent by all reduce tasks (ms)=10716
Total vcore-milliseconds taken by all map tasks=10915
Total vcore-milliseconds taken by all reduce tasks=10716
Total megabyte-milliseconds taken by all map tasks=11176960
```

```
Hadoop 2.6.1.1 [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help

Total vcore-milliseconds taken by all reduce tasks=10716
Total megabyte-milliseconds taken by all map tasks=11176960
Total megabyte-milliseconds taken by all reduce tasks=10973184

Map-Reduce Framework
  Map input records=26
  Map output records=30
  Map output bytes=435
  Map output materialized bytes=39
  Input split bytes=96
  Combine input records=30
  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)=330
  CPU time spent (ms)=2700
  Physical memory (bytes) snapshot=296574976
  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=51
File Output Format Counters
  Bytes Written=19

The mean is: 1.0
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost mapreduce]$
```

```
Hadoop 2.6.1.1 [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help

Map output bytes=435
Map output materialized bytes=39
Input split bytes=96
Combine input records=30
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)=330
CPU time spent (ms)=2700
Physical memory (bytes) snapshot=296574976
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=51
File Output Format Counters
  Bytes Written=19

The mean is: 1.0
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost mapreduce]$ hdfs dfs -cat wordmean/part-r-00000
18/08/26 16:05:01 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
count    15
length   15
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost mapreduce]$
```

Calculate standard deviation:

```
Hadoop 2.6.1.1 [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help

[acadgild@localhost mapreduce]$ hadoop jar hadoop-mapreduce-examples-2.6.5.jar wordstandarddeviation /file1.txt wordstandarddeviation
18/08/26 16:08:35 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/08/26 16:08:37 INFO client.RMProxy: Connecting to ResourceManager at localhost:127.0.0.1:8032
18/08/26 16:08:40 INFO input.FileInputFormat: Total input paths to process : 1
18/08/26 16:08:40 INFO mapreduce.JobSubmitter: number of splits:1
18/08/26 16:08:41 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1535277690582_0003
18/08/26 16:08:41 INFO impl.YarnClientImpl: Submitted application application_1535277690582_0003
18/08/26 16:08:42 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1535277690582_0003/
18/08/26 16:08:42 INFO mapreduce.Job: Running job: job_1535277690582_0003
18/08/26 16:08:59 INFO mapreduce.Job: Job job_1535277690582_0003 running in uber mode : false
18/08/26 16:09:11 INFO mapreduce.Job: map 0% reduce 0%
18/08/26 16:09:24 INFO mapreduce.Job: map 100% reduce 100%
18/08/26 16:09:25 INFO mapreduce.Job: Job job_1535277690582_0003 completed successfully
18/08/26 16:09:26 INFO mapreduce.Job: Counters: 49

File System Counters
  FILE: Number of bytes read=56
  FILE: Number of bytes written=215633
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=147
  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)=10151
  Total time spent by all reduces in occupied slots (ms)=10027
  Total time spent by all map tasks (ms)=10151
  Total time spent by all reduce tasks (ms)=10027
  Total vcore-milliseconds taken by all map tasks=10151
  Total vcore-milliseconds taken by all reduce tasks=10027

acadgild@localhost:~$
```

```
Hadoop 2.6.1.1 [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help

Total vcore-milliseconds taken by all reduce tasks=10027
Total megabyte-milliseconds taken by all map tasks=10394624
Total megabyte-milliseconds taken by all reduce tasks=10267648

Map-Reduce Framework
  Map input records=26
  Map output records=45
  Map output bytes=660
  Map output materialized bytes=56
  Input split bytes=96
  Combine input records=45
  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)=312
  CPU time spent (ms)=2610
  Physical memory (bytes) snapshot=298688512
  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=51
File Output Format Counters
  Bytes Written=29

The standard deviation is: 0.0
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost mapreduce]$
```

Hadoop 2.6.1.1 [Running] - Oracle VM VirtualBox

File Machine View Input Devices Help

```
Map output materialized bytes=56
Input split bytes=96
Combine input records=45
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)=312
CPU time spent (ms)=2610
Physical memory (bytes) snapshot=298688512
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=51
File Output Format Counters
Bytes Written=29

The standard deviation is: 0.0
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost mapreduce]$ hdfs dfs -cat wordstandarddeviation/part-r-00000
18/08/26 16:10:50 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
count  15
length 15
square 15
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost mapreduce]$
```

acadgild@localhost ~

16:11
26-08-2018