

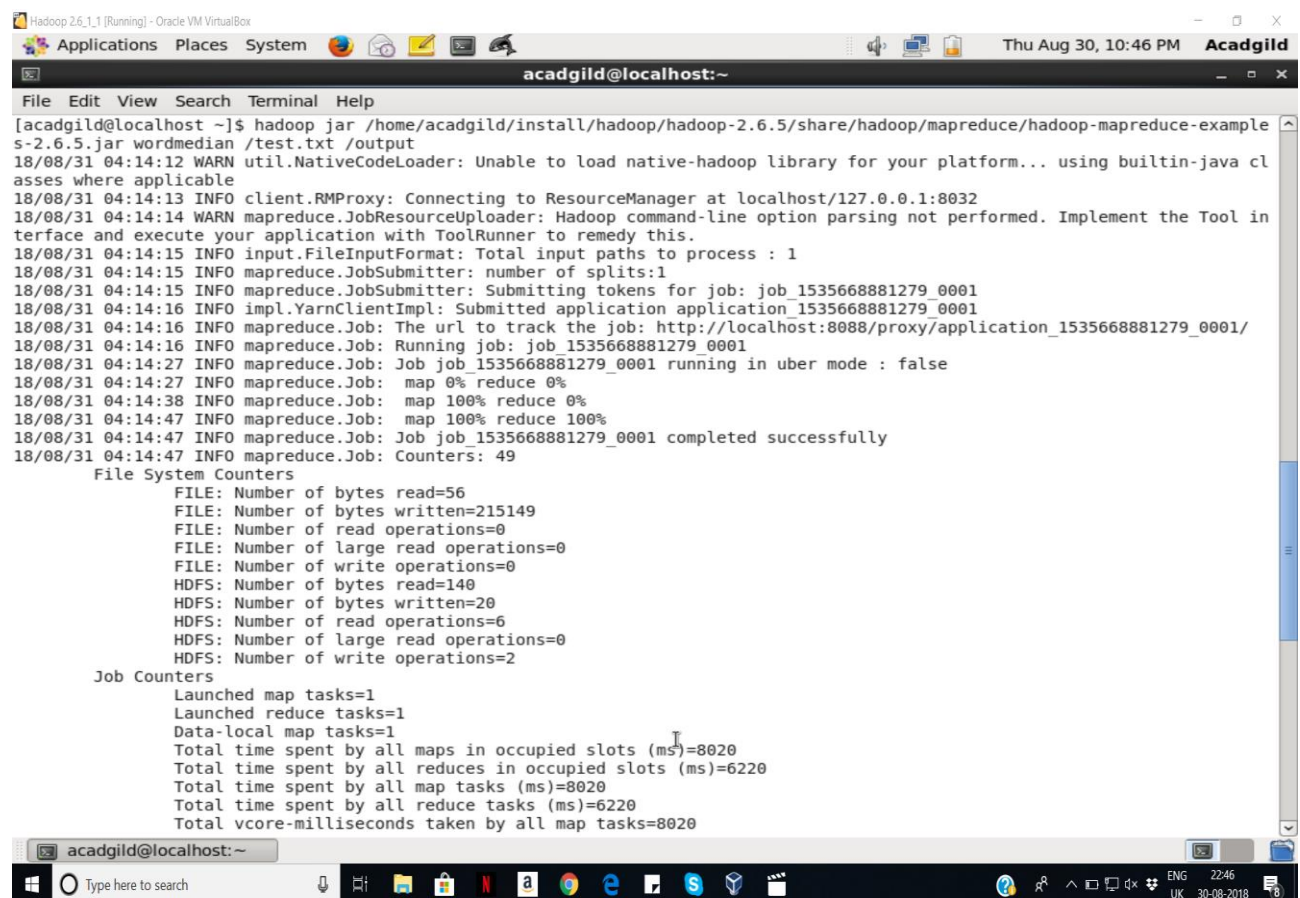
# ASSIGNMENT 3

## TASK 1:

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

## SOLUTION:

### WORDMEDIAN



```
[acadgild@localhost ~]$ hadoop jar /home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce/hadoop-mapreduce-examples-2.6.5.jar wordmedian /test.txt /output
18/08/31 04:14:12 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/08/31 04:14:13 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/08/31 04:14:14 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
18/08/31 04:14:15 INFO input.FileInputFormat: Total input paths to process : 1
18/08/31 04:14:15 INFO mapreduce.JobSubmitter: number of splits:1
18/08/31 04:14:15 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1535668881279_0001
18/08/31 04:14:16 INFO impl.YarnClientImpl: Submitted application application_1535668881279_0001
18/08/31 04:14:16 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1535668881279_0001/
18/08/31 04:14:16 INFO mapreduce.Job: Running job: job_1535668881279_0001
18/08/31 04:14:27 INFO mapreduce.Job: Job job_1535668881279_0001 running in uber mode : false
18/08/31 04:14:27 INFO mapreduce.Job: map 0% reduce 0%
18/08/31 04:14:38 INFO mapreduce.Job: map 100% reduce 0%
18/08/31 04:14:47 INFO mapreduce.Job: map 100% reduce 100%
18/08/31 04:14:47 INFO mapreduce.Job: Job job_1535668881279_0001 completed successfully
18/08/31 04:14:47 INFO mapreduce.Job: Counters: 49
  File System Counters
    FILE: Number of bytes read=56
    FILE: Number of bytes written=215149
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=140
    HDFS: Number of bytes written=20
    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)=8020
    Total time spent by all reduces in occupied slots (ms)=6220
    Total time spent by all map tasks (ms)=8020
    Total time spent by all reduce tasks (ms)=6220
    Total vcore-milliseconds taken by all map tasks=8020
```

```
Hadoop 2.6.1.1 [Running] - Oracle VM VirtualBox
Applications Places System
acadgild@localhost:~
File Edit View Search Terminal Help
Reduce input records=5
Reduce output records=5
Spilled Records=10
Shuffled Maps =1
Failed Shuffles=0
Merged Map outputs=1
GC time elapsed (ms)=168
CPU time spent (ms)=1760
Physical memory (bytes) snapshot=301498368
Virtual memory (bytes) snapshot=4117938176
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=45
File Output Format Counters
Bytes Written=20
The median is: 4
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$ hadoop fs -ls /output
18/08/30 22:47:14 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 2018-08-31 04:14 /output/_SUCCESS
-rw-r--r-- 1 acadgild supergroup 20 2018-08-31 04:14 /output/part-r-00000
[acadgild@localhost ~]$ hadoop fs -cat /output/part-r-00000
18/08/30 22:47:48 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
1      2
2      1
4      4
7      1
9      1
[acadgild@localhost ~]$
```

## COMMANDS USED:

**`_Hadoop jar /home/acadgild/install/hadoop/Hadoop-2.6.5/share/hadoop/mapreduce/Hadoop-mapreduce-examples-2.6.5.jar wordmedian /test.txt /output`**

In this command we are specifying the jar file and path where the wordmedian code or program exists. After specifying the path we type in the input file on which the code is executed that is /test.txt. And then we specify the directory where the output of wordmedian program executed on file test.txt is directed that is /output.

**`_hadoop fs -ls /output`**

This command is used to list all the files present in the output directory. It displays the SUCCESS and PART-R-00000 sub-directories.

**`-hadoop fs -cat /output/part-r-00000`**

This command is used to display the content in part-r-00000.

# WORDMEAN:

```
Hadoop 2.6.1_1 [Running] - Oracle VM VirtualBox
Applications Places System Thu Aug 30, 10:49 PM Acadgild
acadgild@localhost:~
File Edit View Search Terminal Help
[acadgild@localhost ~]$ hadoop jar /home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce/hadoop-mapreduce-exam
s-2.6.5.jar wordmean /test.txt /output
18/08/30 22:48:54 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
18/08/30 22:48:55 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/08/30 22:48:56 INFO input.FileInputFormat: Total input paths to process : 1
18/08/30 22:48:56 INFO mapreduce.JobSubmitter: number of splits:1
18/08/30 22:48:56 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1535668881279_0002
18/08/30 22:48:57 INFO impl.YarnClientImpl: Submitted application application_1535668881279_0002
18/08/30 22:48:57 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1535668881279_0002/
18/08/30 22:48:57 INFO mapreduce.Job: Running job: job_1535668881279_0002
18/08/30 22:49:04 INFO mapreduce.Job: Job job_1535668881279_0002 running in uber mode : false
18/08/30 22:49:04 INFO mapreduce.Job: map 0% reduce 0%
18/08/30 22:49:10 INFO mapreduce.Job: map 100% reduce 0%
18/08/30 22:49:16 INFO mapreduce.Job: map 100% reduce 100%
18/08/30 22:49:17 INFO mapreduce.Job: Job job_1535668881279_0002 completed successfully
18/08/30 22:49:17 INFO mapreduce.Job: Counters: 49
File System Counters
  FILE: Number of bytes read=39
  FILE: Number of bytes written=215379
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=140
  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)=3251
  Total time spent by all reduces in occupied slots (ms)=3939
  Total time spent by all map tasks (ms)=3251
  Total time spent by all reduce tasks (ms)=3939
  Total vcore-milliseconds taken by all map tasks=3251
  Total vcore-milliseconds taken by all reduce tasks=3939
  Total megabyte-milliseconds taken by all map tasks=3329024
```

```
Hadoop 2.6.1_1 [Running] - Oracle VM VirtualBox
Applications Places System Thu Aug 30, 10:50 PM Acadgild
acadgild@localhost:~
File Edit View Search Terminal Help
Combine input records=18
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)=105
CPU time spent (ms)=1060
Physical memory (bytes) snapshot=298487808
Virtual memory (bytes) snapshot=4117938176
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=45
File Output Format Counters
  Bytes Written=18
The mean is: 4.0
[acadgild@localhost ~]$ hadoop fs -ls /output
18/08/30 22:49:57 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 2018-08-30 22:49 /output/_SUCCESS
-rw-r--r-- 1 acadgild supergroup 18 2018-08-30 22:49 /output/part-r-00000
[acadgild@localhost ~]$ hadoop fs -cat /output/part-r-00000
18/08/30 22:50:05 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
count 9
length 36
[acadgild@localhost ~]$
```

## **COMMANDS USED:**

**\_Hadoop jar /home/acadgild/install/hadoop/Hadoop-2.6.5/share/hadoop/mapreduce/Hadoop-mapreduce-examples-2.6.5.jar wordmean /test.txt /output**

In this command we are specifying the jar file and path where the wordmean code or program exists. After specifying the path we type in the input file on which the code is executed that is /test.txt. And then we specify the directory where the output of wordmean program executed on file test.txt is directed that is /output.

**\_hadoop fs -ls /output**

This command is used to list all the files present in the output directory. It displays the SUCCESS and PART-R-00000 sub-directories.

**-hadoop fs -cat /output/part-r-00000**

This command is used to display the content in part-r-00000.

# WORD STANDARD DEVIATION:

```
Hadoop 2.6.1_1 [Running] - Oracle VM VirtualBox
Applications Places System Thu Aug 30, 10:51 PM Acadgild
acadgild@localhost:~
File Edit View Search Terminal Help

[acadgild@localhost ~]$ hadoop jar /home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce/hadoop-mapreduce-exam
s-2.6.5.jar wordstandarddeviation /test.txt /output
18/08/30 22:50:51 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
18/08/30 22:50:51 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/08/30 22:50:53 INFO input.FileInputFormat: Total input paths to process : 1
18/08/30 22:50:53 INFO mapreduce.JobSubmitter: number of splits:1
18/08/30 22:50:53 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1535668881279_0003
18/08/30 22:50:53 INFO impl.YarnClientImpl: Submitted application application_1535668881279_0003
18/08/30 22:50:53 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1535668881279_0003/
18/08/30 22:50:53 INFO mapreduce.Job: Running job: job_1535668881279_0003
18/08/30 22:51:01 INFO mapreduce.Job: Job job_1535668881279_0003 running in uber mode : false
18/08/30 22:51:01 INFO mapreduce.Job: map 0% reduce 0%
18/08/30 22:51:07 INFO mapreduce.Job: map 100% reduce 0%
18/08/30 22:51:13 INFO mapreduce.Job: map 100% reduce 100%
18/08/30 22:51:14 INFO mapreduce.Job: Job job_1535668881279_0003 completed successfully
18/08/30 22:51:14 INFO mapreduce.Job: Counters: 49
  File System Counters
    FILE: Number of bytes read=56
    FILE: Number of bytes written=215573
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=140
    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)=3233
    Total time spent by all reduces in occupied slots (ms)=3903
    Total time spent by all map tasks (ms)=3233
    Total time spent by all reduce tasks (ms)=3903
    Total vcore-milliseconds taken by all map tasks=3233
    Total vcore-milliseconds taken by all reduce tasks=3903

acadgild@localhost:~
Type here to search
Hadoop 2.6.1_1 [Running] - Oracle VM VirtualBox
Applications Places System Thu Aug 30, 10:52 PM Acadgild
acadgild@localhost:~
File Edit View Search Terminal Help
  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)=110
  CPU time spent (ms)=1070
  Physical memory (bytes) snapshot=298057728
  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=45
  File Output Format Counters
    Bytes Written=29
The standard deviation is: 2.494438257849294
[acadgild@localhost ~]$ hadoop fs -ls /output
18/08/30 22:51:57 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 2018-08-30 22:51 /output/_SUCCESS
-rw-r--r-- 1 acadgild supergroup 29 2018-08-30 22:51 /output/part-r-000000
[acadgild@localhost ~]$ hadoop fs -cat /output/part-r-000000
18/08/30 22:52:04 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
count 9
length 36
square 200
[acadgild@localhost ~]$
```



## **COMMANDS USED:**

**\_Hadoop jar /home/acadgild/install/hadoop/Hadoop-2.6.5/share/hadoop/mapreduce/Hadoop-mapreduce-examples-2.6.5.jar wordstandarddeviation /test.txt /output**

In this command we are specifying the jar file and path where the wordstandarddeviation code or program exists. After specifying the path we type in the input file on which the code is executed that is /test.txt. And then we specify the directory where the output of wordstandarddeviationprogram executed on file test.txt is directed that is /output.

**\_hadoop fs -ls /output**

This command is used to list all the files present in the output directory. It displays the SUCCESS and PART-R-00000 sub-directories.

**-hadoop fs -cat /output/part-r-00000**

This command is used to display the content in part-r-00000.

**File used : ( test.txt)**

