Assignment Submission- Session 3

Task 1: Execute **WordMedian** , **WordMean** , **WordStandardDeviation** programs using hadoop-mapreduce-examples-2.9.0.jar file present in your AcadGild VM.

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

Command used is as in screenshot.

```
wordstandarddeviation: A map/reduce program that counts the standard deviation of the length of the words in the input files.
[acadgild@localhost ~]$ clear
[acadgild@localhost ~]$ hadoop jar /home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce/hadoop-mapreduce-examples-2.6.5.jar wordmean /assign3_test.txt /meanOutput
18/08/24 21:04:02 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/08/24 21:04:03 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/08/24 21:04:06 INFO input.FileInputFormat: Total input paths to process : 1
18/08/24 21:04:06 INFO mapreduce.JobSubmitter: number of splits:1
18/08/24 21:04:06 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1535143529220_0001
18/08/24 21:04:07 INFO impl.YarnClientImpl: Submitted application application_1535143529220_0001
18/08/24 21:04:07 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1535143529220_0001/
18/08/24 21:04:07 INFO mapreduce.Job: Running job: job_1535143529220_0001
18/08/24 21:04:20 INFO mapreduce.Job: Job job_1535143529220_0001 running in uber mode : false
18/08/24 21:04:20 INFO mapreduce.Job:  map 0% reduce 0%
18/08/24 21:04:28 INFO mapreduce.Job:  map 100% reduce 0%
18/08/24 21:04:37 INFO mapreduce.Job:  map 100% reduce 100%
18/08/24 21:04:38 INFO mapreduce.Job: Job job_1535143529220_0001 completed successfully
18/08/24 21:04:38 INFO mapreduce.Job: Counters: 49
        File System Counters
                FILE: Number of bytes read=39
                FILE: Number of bytes written=215403
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
```

The output that came on the console: The mean is 5.241379....
Upon using -cat on the output file:
- count of words 58
- total length: 304

```
               Bytes Written=20
The mean is: 5.241379310344827
[acadgild@localhost ~]$ hadoop fs -ls /
18/08/24 21:06:46 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 16 items
drwxr-xr-x   - acadgild supergroup          0 2018-08-24 13:00 /MaxTempTask1Output
drwxr-xr-x   - acadgild supergroup          0 2018-08-24 15:33 /MaxTempTask2_Output
-rw-r--r--   1 acadgild supergroup        217 2018-08-24 12:57 /MaxTemp_DataSet.txt
rwxrwxr--    1 acadgild supergroup          0 2018-08-11 17:19 /aks_test.txt
drwxrwxrwx   - acadgild supergroup          0 2018-08-11 15:46 /akshat
-rw-r--r--   1 acadgild supergroup        365 2018-08-25 02:29 /assign3_test.txt
drwxr-xr-x   - acadgild supergroup          0 2018-08-18 09:38 /hadoopdata
drwxr-xr-x   - acadgild supergroup          0 2018-08-24 21:04 /meanOutput
drwxr-xr-x   - acadgild supergroup          0 2018-08-19 15:17 /myOwnWrittenWC
drwxr-xr-x   - acadgild supergroup          0 2018-08-19 09:05 /myoutput
drwxr-xr-x   - acadgild supergroup          0 2018-08-11 21:11 /sqoopout
drwxr-xr-x   - acadgild supergroup          0 2018-02-02 12:49 /sqoopout111
drwxr-xr-x   - acadgild supergroup          0 2018-08-11 22:15 /sqoopoutbyid
rwxrwxrwx    1 acadgild supergroup        123 2018-08-11 15:28 /test.txt
drwxrwx---   - acadgild supergroup          0 2018-08-20 17:29 /tmp
drwxr-xr-x   - acadgild supergroup          0 2018-08-11 16:05 /user
[acadgild@localhost ~]$ hadoop fs -ls /meanOutput
18/08/24 21:07:10 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 2 items
-rw-r--r--   1 acadgild supergroup          0 2018-08-24 21:04 /meanOutput/_SUCCESS
-rw-r--r--   1 acadgild supergroup         20 2018-08-24 21:04 /meanOutput/part-r-00000
[acadgild@localhost ~]$ hadoop fs -cat /meanOutput/part-r-00000
18/08/24 21:07:33 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
count   58
length  304
[acadgild@localhost ~]$
```

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

Command used is as in screenshot.

```
[acadgild@localhost ~]$ hadoop jar /home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce/hadoop-mapreduce-examples-2.6.5.jar wordmedian /assign3_test.txt /medianOutput
18/08/24 21:19:17 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/08/24 21:19:18 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/08/24 21:19:19 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
18/08/24 21:19:20 INFO input.FileInputFormat: Total input paths to process : 1
18/08/24 21:19:20 INFO mapreduce.JobSubmitter: number of splits:1
18/08/24 21:19:21 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1535143529220_0002
18/08/24 21:19:21 INFO impl.YarnClientImpl: Submitted application application_1535143529220_0002
18/08/24 21:19:21 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1535143529220_0002/
18/08/24 21:19:21 INFO mapreduce.Job: Running job: job_1535143529220_0002
```

The output that came on the console: The median is: 4

```
                Bytes Written=53
The median is: 4
[acadgild@localhost ~]$ hadoop fs -ls /medianOutput
18/08/24 21:20:46 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 2 items
-rw-r--r--   1 acadgild supergroup          0 2018-08-24 21:19 /medianOutput/_SUCCESS
-rw-r--r--   1 acadgild supergroup         53 2018-08-24 21:19 /medianOutput/part-r-00000
[acadgild@localhost ~]$ hadoop fs -cat /medianOutput/part-r-00000
18/08/24 21:21:10 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
1       2
2       3
3       15
4       15
5       2
6       1
7       3
8       9
9       3
10      1
11      3
12      1
[acadgild@localhost ~]$
```

Word Standard Deviation program
A map/reduce program that counts the standard deviation of the length of the words in the input files.

Command used is as in screenshot.

```
    at org.apache.hadoop.util.RunJar.main(RunJar.java:136)
[acadgild@localhost ~]$ hadoop jar /home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce/hadoop-mapreduce-examples-2.6.5.jar wordstandarddeviation /assign3_test.txt /stanDeviationOutput
18/08/24 22:12:02 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/08/24 22:12:03 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/08/24 22:12:05 INFO input.FileInputFormat: Total input paths to process : 1
18/08/24 22:12:05 INFO mapreduce.JobSubmitter: number of splits:1
18/08/24 22:12:05 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1535143529220_0003
18/08/24 22:12:06 INFO impl.YarnClientImpl: Submitted application application_1535143529220_0003
18/08/24 22:12:06 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1535143529220_0003/
```

The output that came on the console: The standard deviation is: 2.7996

Upon using -cat on the output file:
- count of words 58
- total length: 304
- square: 2048

```
                Bytes Written=32
The standard deviation is: 2.7996942248530265
[acadgild@localhost ~]$ hadoop fs -ls /stanDeviationOutput
18/08/24 22:13:16 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 2 items
-rw-r--r--   1 acadgild supergroup          0 2018-08-24 22:12 /stanDeviationOutput/_SUCCESS
-rw-r--r--   1 acadgild supergroup         32 2018-08-24 22:12 /stanDeviationOutput/part-r-00000
[acadgild@localhost ~]$ hadoop fs -cat /stanDeviationOutput/part-r-00000
18/08/24 22:13:40 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
count   58
length  304
square  2048
[acadgild@localhost ~]$
```