Assignment 3.1
YARN

Task 1:

Execute wordmedian , wordmean , wordstandarddeviation programs using

hadoop-mapreduce-examples-2.9.0.jar file present in your AcadGild VM.

Refer path below.

/home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce

1) **Execute wordmedian program :**

This program calculates the median length of words in input file.

We have used command
'**hadoop jar /home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce/hadoop-mapreduce-examples-2.6.5.jar wordmedian /test.txt /wordmedianoutput'**.

Here, **hadoop-mapreduce-examples-2.6.5.jar** is JAR file. **wordmedian** is class name. **/test.txt** is input file path and **/wordmedianoutput** is output directory path.

We get output as :
The median is 5

Assignment 3.1
YARN

```
[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 /wordmedianoutput
18/07/21 22:27:31 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where
applicable
18/07/21 22:27:33 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/07/21 22:27:35 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and
execute your application with ToolRunner to remedy this.
18/07/21 22:27:36 INFO input.FileInputFormat: Total input paths to process : 1
18/07/21 22:27:36 INFO mapreduce.JobSubmitter: number of splits:1
18/07/21 22:27:36 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1532167827900_0019
18/07/21 22:27:37 INFO impl.YarnClientImpl: Submitted application application_1532167827900_0019
18/07/21 22:27:37 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1532167827900_0019/
18/07/21 22:27:37 INFO mapreduce.Job: Running job: job_1532167827900_0019
18/07/21 22:27:53 INFO mapreduce.Job: Job job_1532167827900_0019 running in uber mode : false
18/07/21 22:27:53 INFO mapreduce.Job:  map 0% reduce 0%
18/07/21 22:28:04 INFO mapreduce.Job:  map 100% reduce 0%
18/07/21 22:28:16 INFO mapreduce.Job:  map 100% reduce 100%
18/07/21 22:28:16 INFO mapreduce.Job: Job job_1532167827900_0019 completed successfully
18/07/21 22:28:17 INFO mapreduce.Job: Counters: 49
        File System Counters
                FILE: Number of bytes read=96
                FILE: Number of bytes written=215231
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=254
                HDFS: Number of bytes written=37
                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)=8908
                Total time spent by all reduces in occupied slots (ms)=9338
                Total time spent by all reduce tasks (ms)=9338
                Total vcore-milliseconds taken by all map tasks=8908
                Total vcore-milliseconds taken by all reduce tasks=9338
                Total megabyte-milliseconds taken by all map tasks=9121792
                Total megabyte-milliseconds taken by all reduce tasks=9562112
        Map-Reduce Framework
                Map input records=3
                Map output records=26
                Map output bytes=208
                Map output materialized bytes=96
                Input split bytes=95
                Combine input records=26
                Combine output records=9
                Reduce input groups=9
                Reduce shuffle bytes=96
                Reduce input records=9
                Reduce output records=9
                Spilled Records=18
                Shuffled Maps =1
                Failed Shuffles=0
                Merged Map outputs=1
                GC time elapsed (ms)=236
                CPU time spent (ms)=2660
                Physical memory (bytes) snapshot=300261376
                Virtual memory (bytes) snapshot=4123295744
                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=159
        File Output Format Counters
                Bytes Written=37
The median is: 5
```

In below image you can see that when we see files or directories under Output folder using

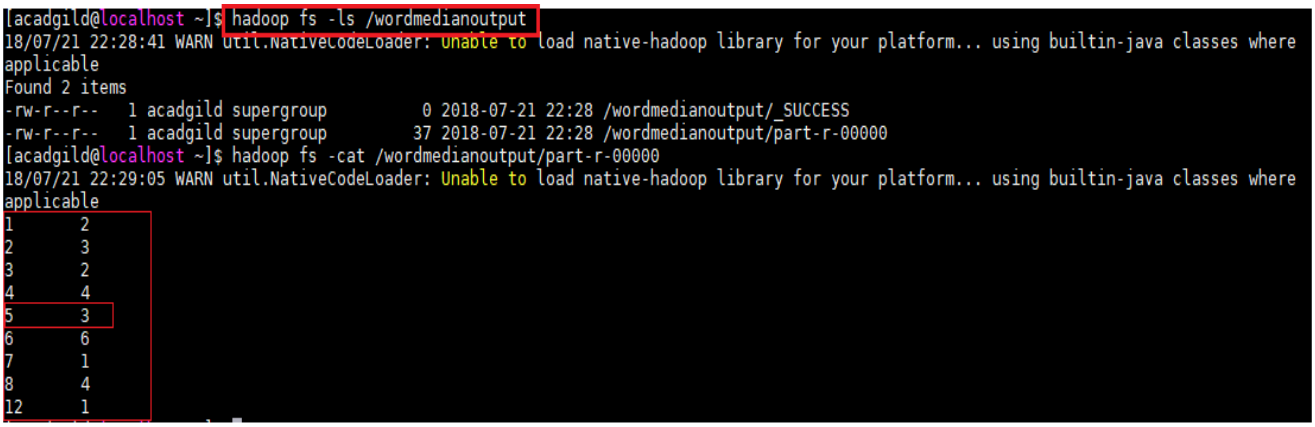
'hadoop fs -ls /wordmedianoutput' command, part-r-00000 file is present.

This file **part-r-00000** contains length of words and their corresponding occurrences.

e.g. 'successfully' word is present in test.txt with only one occurrence. Hence you can see values 12 and 1 in last line as its length is 12 and occurrence is 1.

Here, we can see that there are total 9 values (length of words) in their ascending order of their length and as 5 is coming in the middle (5^{th} element out of total 9). The median is 5.

Assignment 3.1
YARN



2) Execute wordmean program :

This program calculates the average length of words in input file.

We have used command
'hadoop jar /home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce/hadoop-mapreduce-examples-2.6.5.jar wordmean /test.txt /wordmeanoutput'.

Here, **hadoop-mapreduce-examples-2.6.5.jar** is JAR file. **wordmean** is class name. **/test.txt** is input file path and **/wordmeanoutput** is output directory path.

We get output as :
The mean is 5.076923076923077

Assignment 3.1
YARN

```
[acadgild@localhost ~]$ hadoop jar /home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce/hadoop-mapreduce-examples-2.6.5.jar
wordmean /test.txt /wordmeanoutput
18/07/21 22:36:08 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where
applicable
18/07/21 22:36:10 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/07/21 22:36:13 INFO input.FileInputFormat: Total input paths to process : 1
18/07/21 22:36:14 INFO mapreduce.JobSubmitter: number of splits:1
18/07/21 22:36:14 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1532167827900_0020
18/07/21 22:36:15 INFO impl.YarnClientImpl: Submitted application application_1532167827900_0020
18/07/21 22:36:15 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1532167827900_0020/
18/07/21 22:36:15 INFO mapreduce.Job: Running job: job_1532167827900_0020
18/07/21 22:36:31 INFO mapreduce.Job: Job job_1532167827900_0020 running in uber mode : false
18/07/21 22:36:31 INFO mapreduce.Job:  map 0% reduce 0%
18/07/21 22:36:44 INFO mapreduce.Job:  map 100% reduce 0%
18/07/21 22:36:57 INFO mapreduce.Job:  map 100% reduce 100%
18/07/21 22:36:57 INFO mapreduce.Job: Job job_1532167827900_0020 completed successfully
18/07/21 22:36:58 INFO mapreduce.Job: Counters: 49
        File System Counters
                FILE: Number of bytes read=39
                FILE: Number of bytes written=215377
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=254
                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)=9864
                Total time spent by all reduces in occupied slots (ms)=10194
                Total time spent by all map tasks (ms)=9864
                Total time spent by all reduce tasks (ms)=10194
                Total vcore-milliseconds taken by all map tasks=9864
                Total vcore-milliseconds taken by all reduce tasks=10194
                Total megabyte-milliseconds taken by all map tasks=10100736
                Total megabyte-milliseconds taken by all reduce tasks=10438656
        Map-Reduce Framework
                Map input records=3
                Map output records=52
                Map output bytes=754
                Map output materialized bytes=39
                Input split bytes=95
                Combine input records=52
                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)=286
                CPU time spent (ms)=3020
                Physical memory (bytes) snapshot=301727744
                Virtual memory (bytes) snapshot=4118233088
                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=159
        File Output Format Counters
                Bytes Written=20
The mean is: 5.076923076923077
```

In below image you can see that when we see files or directories under Output folder using **'hadoop fs -ls /wordmeanoutput'** command, part-r-00000 file is present.

This file **part-r-00000** contains total count of words and their total length.

Here, total count of words is 26 and their total length is 132.

Hence average length of words is total length/total count of words = (132/26)= 5.076923076923077

```
[acadgild@localhost ~]$ hadoop fs -ls /wordmeanoutput
18/07/21 22:39:50 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 2018-07-21 22:36 /wordmeanoutput/_SUCCESS
-rw-r--r--   1 acadgild supergroup         20 2018-07-21 22:36 /wordmeanoutput/part-r-00000
[acadgild@localhost ~]$ hadoop fs -cat /wordmeanoutput/part-r-00000
18/07/21 22:40:13 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where
applicable
count   26
length  132
```

Assignment 3.1
YARN

3) Execute wordstandarddeviation program :

This program calculates the standard deviation of length of words in input file.

We have used command
'**hadoop jar /home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce/hadoop-mapreduce-examples-2.6.5.jar wordstandarddeviation /test.txt /wordstandarddeviationoutput**'.

Here, **hadoop-mapreduce-examples-2.6.5.jar** is JAR file. **wordstandarddeviation** is class name. **/test.txt** is input file path and **/wordstandarddeviationoutput** is output directory path.

We get output as :
The standard deviation is 2.52560848605877

```
[acadgild@localhost ~]$ hadoop jar /home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce/hadoop-mapreduce-examples-2.6.5.jar
wordstandarddeviation /test.txt /wordstandarddeviationoutput
18/07/21 22:49:32 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where
applicable
18/07/21 22:49:34 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/07/21 22:49:37 INFO input.FileInputFormat: Total input paths to process : 1
18/07/21 22:49:37 INFO mapreduce.JobSubmitter: number of splits:1
18/07/21 22:49:37 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1532167827900_0021
18/07/21 22:49:38 INFO impl.YarnClientImpl: Submitted application application_1532167827900_0021
18/07/21 22:49:38 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1532167827900_0021/
18/07/21 22:49:38 INFO mapreduce.Job: Running job: job_1532167827900_0021
18/07/21 22:49:55 INFO mapreduce.Job: Job job_1532167827900_0021 running in uber mode : false
18/07/21 22:49:55 INFO mapreduce.Job:  map 0% reduce 0%
18/07/21 22:50:07 INFO mapreduce.Job:  map 100% reduce 0%
18/07/21 22:50:21 INFO mapreduce.Job:  map 100% reduce 100%
18/07/21 22:50:22 INFO mapreduce.Job: Job job_1532167827900_0021 completed successfully
18/07/21 22:50:22 INFO mapreduce.Job: Counters: 49
        File System Counters
                FILE: Number of bytes read=56
                FILE: Number of bytes written=215597
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=254
                HDFS: Number of bytes written=31
                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)=9697
                Total time spent by all reduces in occupied slots (ms)=9974
                Total time spent by all map tasks (ms)=9697
                Total time spent by all reduce tasks (ms)=9974
                Total vcore-milliseconds taken by all map tasks=9697
                Total vcore-milliseconds taken by all reduce tasks=9974
                Total megabyte-milliseconds taken by all map tasks=9929728
                Total megabyte-milliseconds taken by all reduce tasks=10213376
        Map-Reduce Framework
                Map input records=3
                Map output records=78
                Map output bytes=1144
                Map output materialized bytes=56
                Input split bytes=95
                Combine input records=78
                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)=261
                CPU time spent (ms)=2710
                Physical memory (bytes) snapshot=296587264
                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=159
        File Output Format Counters
                Bytes Written=31
The standard deviation is: 2.52560848605877
```

Assignment 3.1
YARN

In below image you can see that when we see files or directories under Output folder using

'hadoop fs -ls /wordstandarddeviationoutput' command, part-r-00000 file is present.

This file **part-r-00000** contains total count of words, their total length.

Here, total count of words is 26 and their total length is 132 and square of length of words is 836.

So we are getting Standard deviation as 2.52560848605877.