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

Solution:

First we will see the contents of the file which will be processed with each of the programs by using the hadoop fs -cat command.

```
[acadgild@localhost mapreduce]$ hadoop fs -cat hadoop/word-count.txt
18/12/03 04:01:33 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
Long years ago, we made atryst with destiny and now the time comes when we shall redeem our pledge, not wholly or in full mea
sure but very substantially. At the stroke of the midnight hour, when the world sleeps, India will awake to life and freedom,
 We take the pledge of dedication to the service of India. --- Tryst with Destiny
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost mapreduce]$
```

Now we shall see the processing of various programs.

Command syntax : hadoop jar <jar file> <driver class> <input path in hdfs> <output file path>

a)**WordMedian**

The hadoop jar command is used to run a jar file. The Word Median program reads files from an input directory,performs its job, and writes the results of the job to files in an output directory (medianOutput) .

```
[acadgild@localhost mapreduce]$ hadoop jar hadoop-mapreduce-examples-2.6.5.jar wordmedian hadoop/word-count.txt  medianOutput

18/12/03 03:16:53 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
18/12/03 03:16:55 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/12/03 03:16:58 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool in
terface and execute your application with ToolRunner to remedy this.
18/12/03 03:16:59 INFO input.FileInputFormat: Total input paths to process : 1
18/12/03 03:16:59 INFO mapreduce.JobSubmitter: number of splits:1
18/12/03 03:17:00 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1543783125070_0007
18/12/03 03:17:01 INFO impl.YarnClientImpl: Submitted application application_1543783125070_0007
18/12/03 03:17:01 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1543783125070_0007/
18/12/03 03:17:01 INFO mapreduce.Job: Running job: job_1543783125070_0007
18/12/03 03:17:11 INFO mapreduce.Job: Job job_1543783125070_0007 running in uber mode : false
18/12/03 03:17:11 INFO mapreduce.Job:  map 0% reduce 0%
18/12/03 03:17:31 INFO mapreduce.Job:  map 100% reduce 0%
18/12/03 03:17:48 INFO mapreduce.Job:  map 100% reduce 100%
18/12/03 03:17:49 INFO mapreduce.Job: Job job_1543783125070_0007 completed successfully
18/12/03 03:17:49 INFO mapreduce.Job: Counters: 49
        File System Counters
                FILE: Number of bytes read=96
                FILE: Number of bytes written=215323
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=454
                HDFS: Number of bytes written=41
                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

                Failed Shuffles=0
                Merged Map outputs=1
                GC time elapsed (ms)=296
                CPU time spent (ms)=2920
                Physical memory (bytes) snapshot=294580224
                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=332
        File Output Format Counters
                Bytes Written=41
The median is: 4
```

The hadoop fs -cat command is used to verify the output.

```
[acadgild@localhost mapreduce]$ hadoop fs -ls medianOutput
18/12/03 03:35:55 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 2018-12-03 03:17 medianOutput/_SUCCESS
-rw-r--r--   1 acadgild supergroup         41 2018-12-03 03:17 medianOutput/part-r-00000
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost mapreduce]$ hadoop fs -cat medianOutput/part-r-00000
18/12/03 03:36:22 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
2       11
3       13
4       13
5       8
6       6
7       6
8       2
10      1
14      1
[acadgild@localhost mapreduce]$
```

b) **WordMean :**

The hadoop jar command is used to run a jar file. The Word Mean program reads files from an input directory,performs its job, and writes the results of the job to files in an output directory (meanOutput) .

```
[acadgild@localhost mapreduce]$ hadoop jar hadoop-mapreduce-examples-2.6.5.jar wordmean hadoop/word-count.txt  meanOutput
18/12/03 03:42:13 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
18/12/03 03:42:17 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/12/03 03:42:23 INFO input.FileInputFormat: Total input paths to process : 1
18/12/03 03:42:24 INFO mapreduce.JobSubmitter: number of splits:1
18/12/03 03:42:25 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1543783125070_0008
18/12/03 03:42:26 INFO impl.YarnClientImpl: Submitted application application_1543783125070_0008
18/12/03 03:42:27 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1543783125070_0008/
18/12/03 03:42:27 INFO mapreduce.Job: Running job: job_1543783125070_0008
18/12/03 03:42:59 INFO mapreduce.Job: Job job_1543783125070_0008 running in uber mode : false
18/12/03 03:42:59 INFO mapreduce.Job:  map 0% reduce 0%
18/12/03 03:43:24 INFO mapreduce.Job:  map 100% reduce 0%
18/12/03 03:43:49 INFO mapreduce.Job:  map 100% reduce 100%
18/12/03 03:43:50 INFO mapreduce.Job: Job job_1543783125070_0008 completed successfully
18/12/03 03:43:51 INFO mapreduce.Job: Counters: 49
        File System Counters
                FILE: Number of bytes read=39
                FILE: Number of bytes written=215469
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=454
                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)=20291
                Total time spent by all reduces in occupied slots (ms)=21455
                Total time spent by all map tasks (ms)=20291
                Total time spent by all reduce tasks (ms)=21455
                Total vcore-milliseconds taken by all map tasks=20291
```

```
                Map output materialized bytes=39
                Input split bytes=122
                Combine input records=122
                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)=573
                CPU time spent (ms)=5420
                Physical memory (bytes) snapshot=295620608
                Virtual memory (bytes) snapshot=4126769152
                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=332
        File Output Format Counters
                Bytes Written=20
The mean is: 4.4426229508196721
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost mapreduce]$ hadoop fs -cat meanOutput/part-r-00000
18/12/03 03:47:55 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
count   61
length  271
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost mapreduce]$
```

c) **WordStandardDeviation :**

The hadoop jar command is used to run a jar file. The Word Standard Deviation program reads files from an input directory,performs its job, and writes the results of the job to files in an output directory (standardDeviationOutput) .

```
[acadgild@localhost mapreduce]$ hadoop jar hadoop-mapreduce-examples-2.6.5.jar wordstandarddeviation hadoop/word-count.txt  s
tandardDeviationOutput
18/12/03 03:54:37 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
18/12/03 03:54:41 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/12/03 03:54:47 INFO input.FileInputFormat: Total input paths to process : 1
18/12/03 03:54:47 INFO mapreduce.JobSubmitter: number of splits:1
18/12/03 03:54:48 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1543783125070_0009
18/12/03 03:54:49 INFO impl.YarnClientImpl: Submitted application application_1543783125070_0009
18/12/03 03:54:49 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1543783125070_0009/
18/12/03 03:54:49 INFO mapreduce.Job: Running job: job_1543783125070_0009
18/12/03 03:55:18 INFO mapreduce.Job: Job job_1543783125070_0009 running in uber mode : false
18/12/03 03:55:18 INFO mapreduce.Job:  map 0% reduce 0%
18/12/03 03:55:39 INFO mapreduce.Job:  map 100% reduce 0%
18/12/03 03:56:01 INFO mapreduce.Job:  map 100% reduce 100%
18/12/03 03:56:03 INFO mapreduce.Job: Job job_1543783125070_0009 completed successfully
18/12/03 03:56:04 INFO mapreduce.Job: Counters: 49
        File System Counters
                FILE: Number of bytes read=56
                FILE: Number of bytes written=215689
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=454
                HDFS: Number of bytes written=32
                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

                WRONG_REDUCE=0
        File Input Format Counters
                Bytes Read=332
        File Output Format Counters
                Bytes Written=32
The standard deviation is: 2.2141466932309144
```

```
[acadgild@localhost mapreduce]$ hadoop fs -cat standardDeviationOutput/part-r-00000
18/12/03 03:58:52 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
count   61
length  271
square  1503
[acadgild@localhost mapreduce]$
```