Task 1: Execute Word Mean program using hadoop mapreduce example
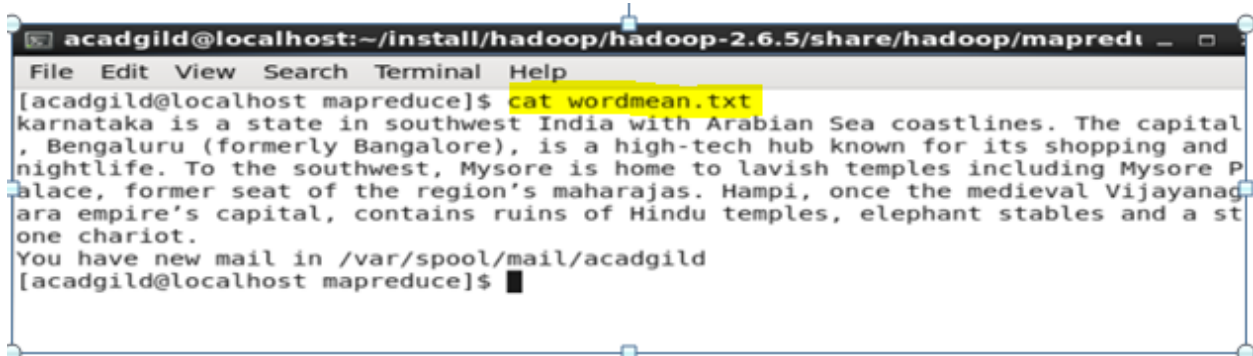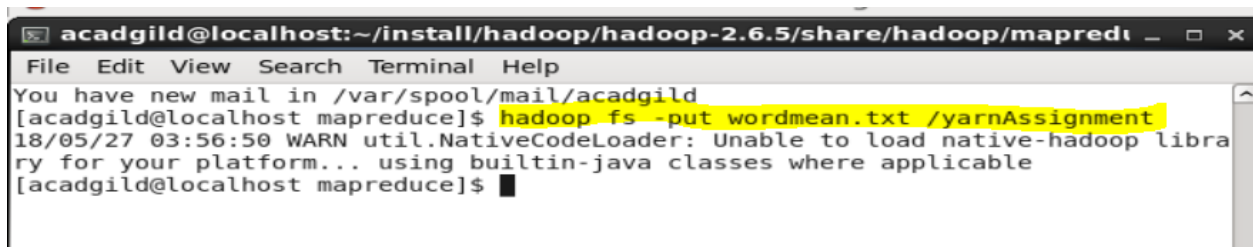
    a. Created wordmean.txt in local file system

    b. Copied wordmean.txt file into hadoop fs in /yarnAssignment directory

    c. Executing **wordmean.txt** program using hadoop-mapreduce-example-2.9.0.jar and coping executed file into **/yarnAssignment/Output** Directory

```
[acadgild@localhost mapreduce]$ hadoop jar hadoop-mapreduce-examples-2.6.5.jar w
ordmean /yarnAssignment/wordmean.txt /yarnAssignment/output
18/05/27 04:08:29 WARN util.NativeCodeLoader: Unable to load native-hadoop libra
ry for your platform... using builtin-java classes where applicable
18/05/27 04:08:31 INFO client.RMProxy: Connecting to ResourceManager at localhos
t/127.0.0.1:8032
18/05/27 04:08:36 INFO input.FileInputFormat: Total input paths to process : 1
18/05/27 04:08:36 INFO mapreduce.JobSubmitter: number of splits:1
18/05/27 04:08:37 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_15
27365639903_0001
18/05/27 04:08:39 INFO impl.YarnClientImpl: Submitted application application_15
27365639903_0001
18/05/27 04:08:39 INFO mapreduce.Job: The url to track the job: http://localhost
:8088/proxy/application_1527365639903_0001/
18/05/27 04:08:39 INFO mapreduce.Job: Running job: job_1527365639903_0001
18/05/27 04:09:08 INFO mapreduce.Job: Job job_1527365639903_0001 running in uber
 mode : false
18/05/27 04:09:08 INFO mapreduce.Job:  map 0% reduce 0%
18/05/27 04:09:24 INFO mapreduce.Job:  map 100% reduce 0%
18/05/27 04:09:44 INFO mapreduce.Job:  map 100% reduce 100%
18/05/27 04:09:45 INFO mapreduce.Job: Job job_1527365639903_0001 completed succe
ssfully
18/05/27 04:09:46 INFO mapreduce.Job: Counters: 49
        File System Counters
                FILE: Number of bytes read=39
                FILE: Number of bytes written=215447
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=531
                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)=12652
                Total time spent by all reduces in occupied slots (ms)=17535
                Total time spent by all map tasks (ms)=12652
                Total time spent by all reduce tasks (ms)=17535
                Total vcore-milliseconds taken by all map tasks=12652
                Total vcore-milliseconds taken by all reduce tasks=17535
                Total megabyte-milliseconds taken by all map tasks=12955648
                Total megabyte-milliseconds taken by all reduce tasks=17955840
        Map-Reduce Framework
```

d. The mean value of **wordmean.txt** file.

```
                                    ....  ..........  .... ...... .., ..., ....  .....
        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=417
        File Output Format Counters
                Bytes Written=20
The mean is: 5.661290322580645
```

e. localhost:50070 – UI representation

/yarnAssignment/output							Go!

Permission	Owner	Group	Size	Replication	Block Size	Name
-rw-r--r--	acadgild	supergroup	0 B	1	128 MB	_SUCCESS
-rw-r--r--	acadgild	supergroup	20 B	1	128 MB	part-r-00000

Task 2: Execute Word Median program using hadoop mapreduce example

a.   Created **wordmedian.txt** in local file system

```
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost mapreduce]$ vi wordmedian.txt
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost mapreduce]$ cat wordmedian.txt
India is a vast South Asian country with diverse terrain — from Himalayan peaks to Indian Ocean coastline — and history rea
ching back 5 millennia. In the north, Mughal Empire landmarks include Delhi's Red Fort complex and massive Jama Masjid mosq
ue, plus Agra's iconic Taj Mahal mausoleum. Pilgrims bathe in the Ganges in Varanasi, and Rishikesh is a yoga centre and ba
se for Himalayan trekking
[acadgild@localhost mapreduce]$ █
```

b.   Copied **wordmedian.txt** file into hadoop fs in /yarnAssignment directory

c.   Executing **wordmean.txt** program using hadoop-mapreduce-example-2.9.0.jar and coping executed file into /yarnAssignment/**OutputMedian** Directory

```
[acadgild@localhost mapreduce]$ hadoop fs -put wordmedian.txt /yarnAssignment
18/05/27 04:16:57 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java
classes where applicable
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost mapreduce]$ hadoop jar hadoop-mapreduce-examples-2.6.5.jar wordmedian /yarnAssignment/wordmedian.txt /y
arnAssignment/outputMedian
18/05/27 04:17:42 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java
classes where applicable
18/05/27 04:17:45 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/05/27 04:17:48 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool
interface and execute your application with ToolRunner to remedy this.
18/05/27 04:17:49 INFO input.FileInputFormat: Total input paths to process : 1
18/05/27 04:17:49 INFO mapreduce.JobSubmitter: number of splits:1
18/05/27 04:17:50 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1527365639903_0002
18/05/27 04:17:51 INFO impl.YarnClientImpl: Submitted application application_1527365639903_0002
18/05/27 04:17:51 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1527365639903_0002/
18/05/27 04:17:51 INFO mapreduce.Job: Running job: job_1527365639903_0002
18/05/27 04:18:13 INFO mapreduce.Job: Job job_1527365639903_0002 running in uber mode : false
18/05/27 04:18:13 INFO mapreduce.Job:  map 0% reduce 0%
18/05/27 04:18:26 INFO mapreduce.Job:  map 100% reduce 0%
18/05/27 04:18:44 INFO mapreduce.Job:  map 100% reduce 100%
18/05/27 04:18:45 INFO mapreduce.Job: Job job_1527365639903_0002 completed successfully
18/05/27 04:18:45 INFO mapreduce.Job: Counters: 49
        File System Counters
                FILE: Number of bytes read=106
                FILE: Number of bytes written=215333
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=519
                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
                Data-local map tasks=1
                Total time spent by all maps in occupied slots (ms)=11218
                Total time spent by all reduces in occupied slots (ms)=13942
                Total time spent by all map tasks (ms)=11218
                Total time spent by all reduce tasks (ms)=13942
                Total vcore-milliseconds taken by all map tasks=11218
                Total vcore-milliseconds taken by all reduce tasks=13942
                Total megabyte-milliseconds taken by all map tasks=11487232
                Total megabyte-milliseconds taken by all reduce tasks=14276608
```

d.   The median value of wordmedian.txt file.

```
                Total megabyte-milliseconds taken by all reduce tasks=14270
        Map-Reduce Framework
                Map input records=1
                Map output records=65
                Map output bytes=520
                Map output materialized bytes=106
                Input split bytes=116
                Combine input records=65
                Combine output records=10
                Reduce input groups=10
                Reduce shuffle bytes=106
                Reduce input records=10
                Reduce output records=10
                Spilled Records=20
                Shuffled Maps =1
                Failed Shuffles=0
                Merged Map outputs=1
                GC time elapsed (ms)=266
                CPU time spent (ms)=3130
                Physical memory (bytes) snapshot=291549184
                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=403
        File Output Format Counters
                Bytes Written=41
The median is: 5
```

e.   localhost:50070 – UI representation

/yarnAssignment/outputMedian							Go!
Permission	**Owner**	**Group**	**Size**	**Replication**	**Block Size**	**Name**	
-rw-r--r--	acadgild	supergroup	0 B	1	128 MB	_SUCCESS	
-rw-r--r--	acadgild	supergroup	41 B	1	128 MB	part-r-00000	

Task 3: Execute Word Standard Deviation program using hadoop mapreduce example

a.   Created **wordStandardDeviation.txt** in local file system

b.   Copied **wordStandardDeviation.txt** file into hadoop fs in /yarnAssignment directory

c.   Executing wordmean.txt program using hadoop-mapreduce-example-2.9.0.jar and coping executed file into /yarnAssignment/**OutputStandardDeviation** Directory

```
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost mapreduce]$ vi wordStandardDeviation.txt
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost mapreduce]$ cat wordStandardDeviation.txt
Bengaluru (also called Bangalore) is the capital of India's southern Karnataka state. The center of India's high-tech indus
try, the city is also known for its parks and nightlife. By Cubbon Park, Vidhana Soudha is a Neo-Dravidian legislative buil
ding. Former royal residences include 19th-century Bangalore Palace, modeled after England's Windsor Castle, and Tipu Sulta
n's Summer Palace, an 18th-century teak structure.
[acadgild@localhost mapreduce]$ hadoop fs -put wordStandardDeviation.txt /yarnAssignment
18/05/27 04:24:23 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java
classes where applicable
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost mapreduce]$ hadoop jar hadoop-mapreduce-examples-2.6.5.jar wordStandardDeviation /yarnAssignment/wordSt
andardDeviation.txt /yarnAssignment/outputStandardDeviation
```

d.   The mean value of **wordStandardDeviation.txt** file.

```
                  Total megabyte-milliseconds taken by all reduce tasks=16420864
        Map-Reduce Framework
                Map input records=1
                Map output records=177
                Map output bytes=2596
                Map output materialized bytes=56
                Input split bytes=127
                Combine input records=177
                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)=297
                CPU time spent (ms)=3290
                Physical memory (bytes) snapshot=293990400
                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=424
        File Output Format Counters
                Bytes Written=32
The standard deviation is: 2.9463263294166797
```

e. localhost:50070 – UI representation

/yarnAssignment/outputStandardDeviation   Go!

Permission	Owner	Group	Size	Replication	Block Size	Name
-rw-r--r--	acadgild	supergroup	0 B	1	128 MB	_SUCCESS
-rw-r--r--	acadgild	supergroup	32 B	1	128 MB	part-r-00000

The Final output files of all 3 tasks.

Permission	Owner	Group	Size	Replication	Block Size	Name
drwxr-xr-x	acadgild	supergroup	0 B	0	0 B	output
drwxr-xr-x	acadgild	supergroup	0 B	0	0 B	outputMedian
drwxr-xr-x	acadgild	supergroup	0 B	0	0 B	outputStandardDeviation
-rw-r--r--	acadgild	supergroup	424 B	1	128 MB	wordStandardDeviation.txt
-rw-r--r--	acadgild	supergroup	417 B	1	128 MB	wordmean.txt
-rw-r--r--	acadgild	supergroup	403 B	1	128 MB	wordmedian.txt