

Problem Statement

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

- Create a text File containing N Number of Words in local File System.

Command: Vi wordcount.txt

- Put the file to hdfs File System

Command:

Hdfs dfs -put wordcount.txt /user/acadgiid/wordcount.txt

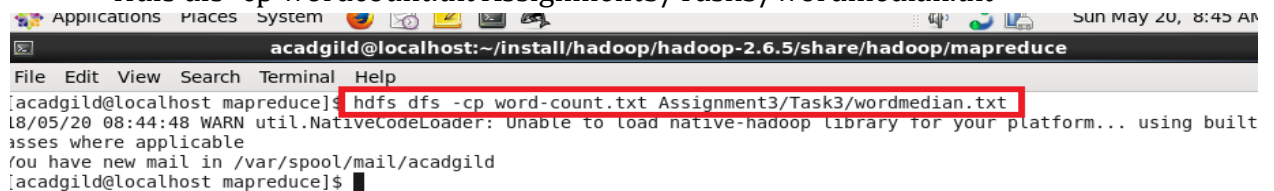
1. Word Median

Step1:

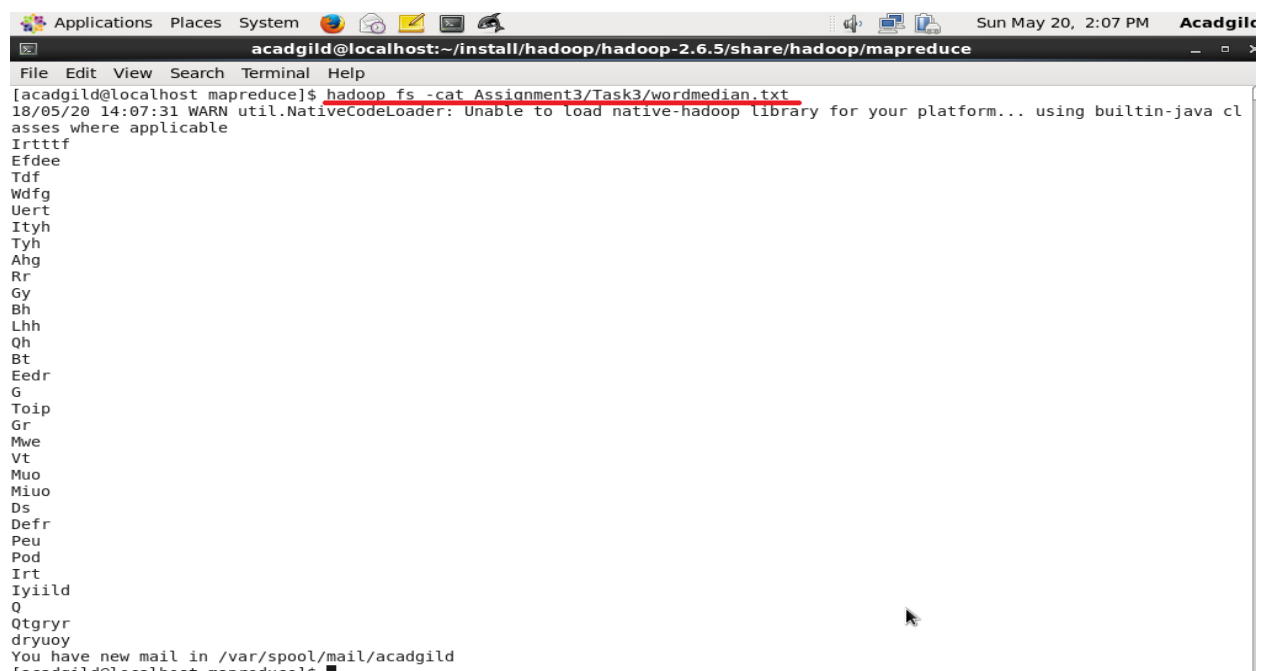
Copy the file wordcount.txt to wordmedian.txt

Command:

Hdfs dfs -cp wordcount.txt Assignment3/Task3/wordmedian.txt



```
acadmild@localhost: ~/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce
[acadmild@localhost mapreduce]$ hdfs dfs -cp word-count.txt Assignment3/Task3/wordmedian.txt
18/05/20 08:44:48 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using built
asses where applicable
You have new mail in /var/spool/mail/acadmild
[acadmild@localhost mapreduce]$
```



```
acadmild@localhost: ~/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce
[acadmild@localhost mapreduce]$ hadoop fs -cat Assignment3/Task3/wordmedian.txt
18/05/20 14:07:31 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
Irtttf
Efdee
Tdf
Wdfg
Uert
Ityh
Tyh
Ahg
Rr
Gy
Bh
Lhh
Qh
Bt
Eedr
G
Toip
Gr
Mwe
Vt
Muoo
Miuo
Ds
Defr
Peu
Pod
Irt
Iyild
Q
Qtgryr
dryuoy
You have new mail in /var/spool/mail/acadmild
[acadmild@localhost mapreduce]$
```

Step2:

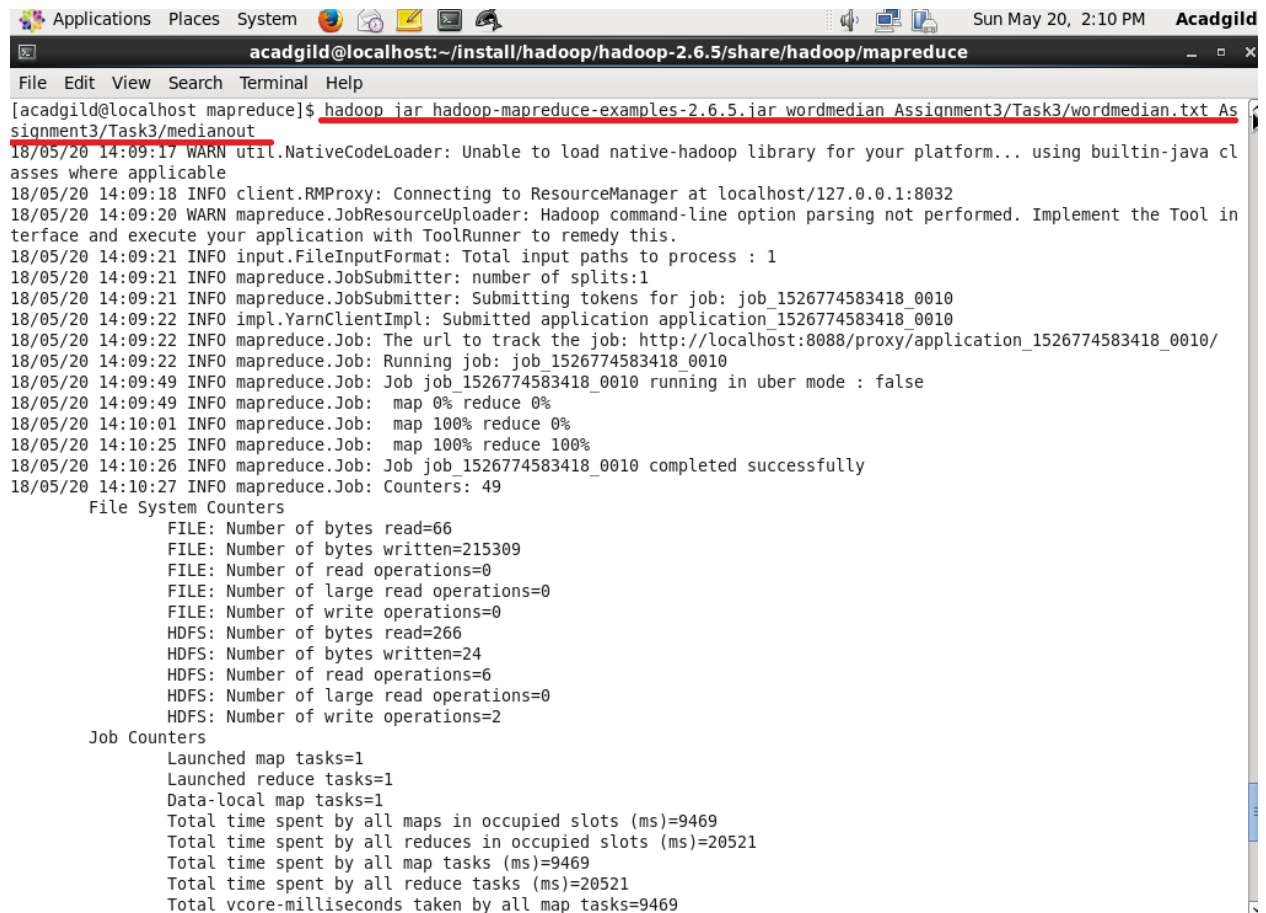
Execute the jar file for word median.

Syntax:

Hadoop jar <jarfilename> <mainclass name> <inputfile path>
<outputFilePath>

Command:

Hadoop jar Hadoop-mapreduce-examples-2.5.6.jar wordmedian
Assignment3/Task3/wordmedian.txt Assignment3/Task3/medianout



```
acadgild@localhost: ~/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce
File Edit View Search Terminal Help
[acadgild@localhost mapreduce]$ hadoop jar hadoop-mapreduce-examples-2.6.5.jar wordmedian Assignment3/Task3/wordmedian.txt Assignment3/Task3/medianout
18/05/20 14:09:17 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/05/20 14:09:18 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/05/20 14:09:20 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
18/05/20 14:09:21 INFO input.FileInputFormat: Total input paths to process : 1
18/05/20 14:09:21 INFO mapreduce.JobSubmitter: number of splits:1
18/05/20 14:09:21 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1526774583418_0010
18/05/20 14:09:22 INFO impl.YarnClientImpl: Submitted application application_1526774583418_0010
18/05/20 14:09:22 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1526774583418_0010/
18/05/20 14:09:22 INFO mapreduce.Job: Running job: job_1526774583418_0010
18/05/20 14:09:49 INFO mapreduce.Job: Job job_1526774583418_0010 running in uber mode : false
18/05/20 14:09:49 INFO mapreduce.Job: map 0% reduce 0%
18/05/20 14:10:01 INFO mapreduce.Job: map 100% reduce 0%
18/05/20 14:10:25 INFO mapreduce.Job: map 100% reduce 100%
18/05/20 14:10:26 INFO mapreduce.Job: Job job_1526774583418_0010 completed successfully
18/05/20 14:10:27 INFO mapreduce.Job: Counters: 49
  File System Counters
    FILE: Number of bytes read=66
    FILE: Number of bytes written=215309
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=266
    HDFS: Number of bytes written=24
    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)=9469
    Total time spent by all reduces in occupied slots (ms)=20521
    Total time spent by all map tasks (ms)=9469
    Total time spent by all reduce tasks (ms)=20521
    Total vcore-milliseconds taken by all map tasks=9469
```

```
Applications Places System Sun May 20, 2:11 PM Acadgild
acadgild@localhost: ~/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce
File Edit View Search Terminal Help
Total vcore-milliseconds taken by all map tasks=9469
Total vcore-milliseconds taken by all reduce tasks=20521
Total megabyte-milliseconds taken by all map tasks=9696256
Total megabyte-milliseconds taken by all reduce tasks=21013504
Map-Reduce Framework
  Map input records=31
  Map output records=31
  Map output bytes=248
  Map output materialized bytes=66
  Input split bytes=133
  Combine input records=31
  Combine output records=6
  Reduce input groups=6
  Reduce shuffle bytes=66
  Reduce input records=6
  Reduce output records=6
  Spilled Records=12
  Shuffled Maps =1
  Failed Shuffles=0
  Merged Map outputs=1
  GC time elapsed (ms)=458
  CPU time spent (ms)=4340
  Physical memory (bytes) snapshot=297959424
  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=133
File Output Format Counters
  Bytes Written=24
The median is: 3
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost mapreduce]$
```

Step 3:

Display the contents of output File

```
Applications Places System Sun May 20, 2:13 PM Acadgild
acadgild@localhost: ~/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce
File Edit View Search Terminal Help
[acadgild@localhost mapreduce]$ hadoop fs -ls Assignment3/Task3/medianout
18/05/20 14:12:30 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 2018-05-20 14:10 Assignment3/Task3/medianout/ SUCCESS
-rw-r--r-- 1 acadgild supergroup        24 2018-05-20 14:10 Assignment3/Task3/medianout/ part-r-00000
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost mapreduce]$ hadoop fs -cat Assignment3/Task3/medianout/part-r-00000
18/05/20 14:13:06 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
1      2
2      8
3      9
4      7
5      1
6      4
[acadgild@localhost mapreduce]$
```

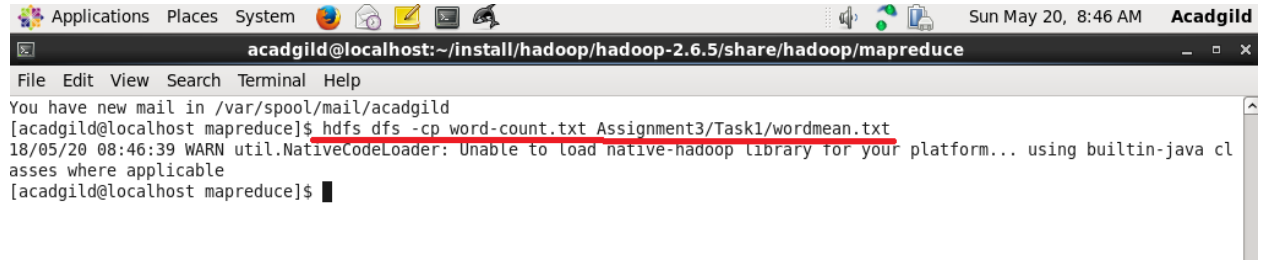
2. Word Mean

Step1:

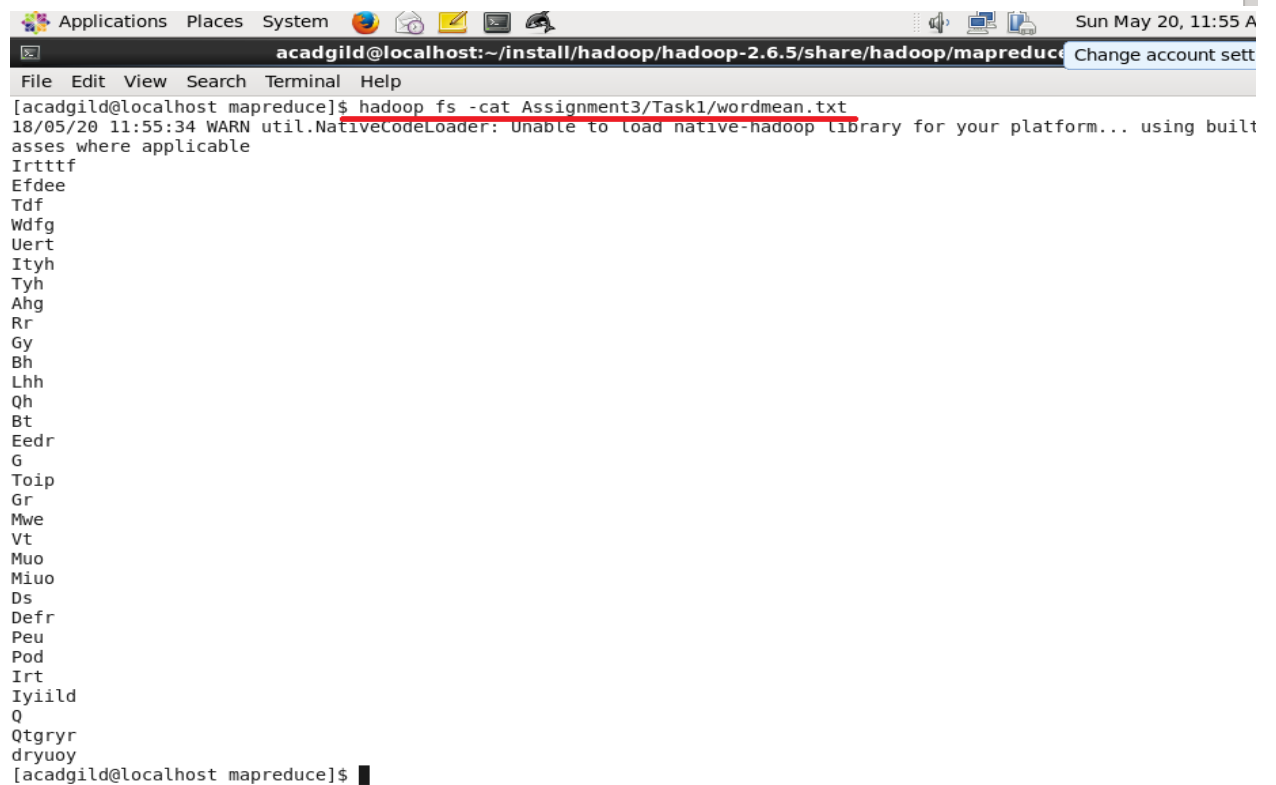
Copy the file wordcount.txt to wordmean.txt

Command:

Hdfs dfs -cp wordcount.txt Assignment3/Task1/wordmean.txt



A terminal window titled 'acadgild@localhost: ~/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce'. The terminal shows the command 'hdfs dfs -cp word-count.txt Assignment3/Task1/wordmean.txt' being executed. A warning message is displayed: '18/05/20 08:46:39 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable'. The prompt returns to '[acadgild@localhost mapreduce]\$'.



A terminal window titled 'acadgild@localhost: ~/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce'. The terminal shows the command 'hadoop fs -cat Assignment3/Task1/wordmean.txt' being executed. A warning message is displayed: '18/05/20 11:55:34 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable'. The output of the command is displayed as a list of words: 'Irtttf', 'Efdee', 'Tdf', 'Wdfg', 'Uert', 'Ityh', 'Tyh', 'Ahg', 'Rr', 'Gy', 'Bh', 'Lhh', 'Qh', 'Bt', 'Eedr', 'G', 'Toip', 'Gr', 'Mwe', 'Vt', 'Muo', 'Miuo', 'Ds', 'Defr', 'Peu', 'Pod', 'Irt', 'Iyiild', 'Q', 'Qtgryr', 'dryuoy'. The prompt returns to '[acadgild@localhost mapreduce]\$'.

Step2:

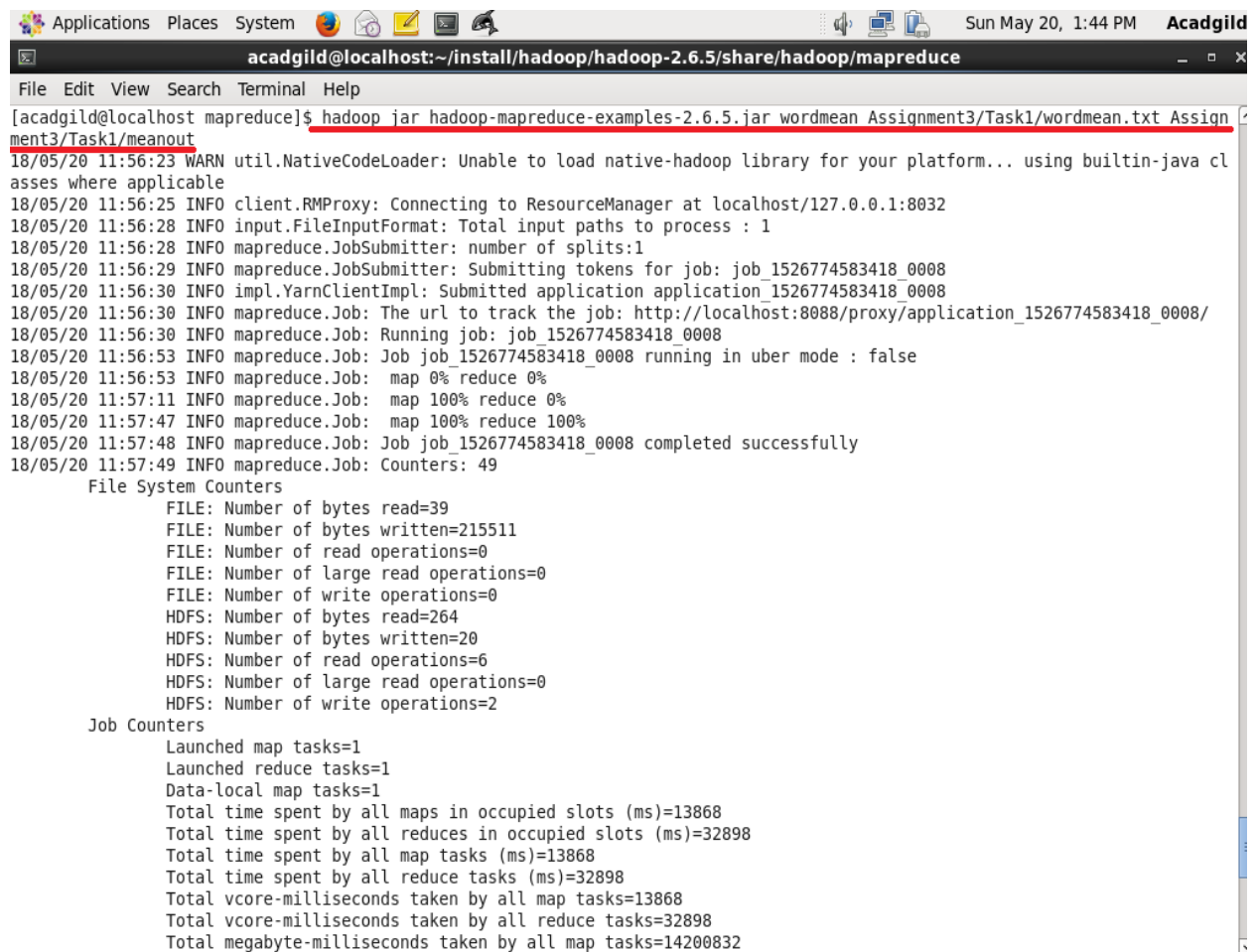
Execute the jar file for word mean.

Syntax:

Hadoop jar <jarfilename> <mainclass name> <inputfile path>
<outputFilePath>

Command:

Hadoop jar Hadoop-mapreduce-examples-2.5.6.jar wordmean
Assignment3/Task1/wordmean.txt Assignment3/Task1/meanout



```
acadgild@localhost:~/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce
File Edit View Search Terminal Help
[acadgild@localhost mapreduce]$ hadoop jar hadoop-mapreduce-examples-2.6.5.jar wordmean Assignment3/Task1/wordmean.txt Assignment3/Task1/meanout
18/05/20 11:56:23 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
18/05/20 11:56:25 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/05/20 11:56:28 INFO input.FileInputFormat: Total input paths to process : 1
18/05/20 11:56:28 INFO mapreduce.JobSubmitter: number of splits:1
18/05/20 11:56:29 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1526774583418_0008
18/05/20 11:56:30 INFO impl.YarnClientImpl: Submitted application application_1526774583418_0008
18/05/20 11:56:30 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1526774583418_0008/
18/05/20 11:56:30 INFO mapreduce.Job: Running job: job_1526774583418_0008
18/05/20 11:56:53 INFO mapreduce.Job: Job job_1526774583418_0008 running in uber mode : false
18/05/20 11:56:53 INFO mapreduce.Job: map 0% reduce 0%
18/05/20 11:57:11 INFO mapreduce.Job: map 100% reduce 0%
18/05/20 11:57:47 INFO mapreduce.Job: map 100% reduce 100%
18/05/20 11:57:48 INFO mapreduce.Job: Job job_1526774583418_0008 completed successfully
18/05/20 11:57:49 INFO mapreduce.Job: Counters: 49
  File System Counters
    FILE: Number of bytes read=39
    FILE: Number of bytes written=215511
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=264
    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)=13868
    Total time spent by all reduces in occupied slots (ms)=32898
    Total time spent by all map tasks (ms)=13868
    Total time spent by all reduce tasks (ms)=32898
    Total vcore-milliseconds taken by all map tasks=13868
    Total vcore-milliseconds taken by all reduce tasks=32898
    Total megabyte-milliseconds taken by all map tasks=14200832
```

```
Applications Places System Sun May 20, 1:44 PM Acadgild
acadgild@localhost:~/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce
File Edit View Search Terminal Help
Total vcore-milliseconds taken by all map tasks=13868
Total vcore-milliseconds taken by all reduce tasks=32898
Total megabyte-milliseconds taken by all map tasks=14200832
Total megabyte-milliseconds taken by all reduce tasks=33687552
Map-Reduce Framework
  Map input records=31
  Map output records=62
  Map output bytes=899
  Map output materialized bytes=39
  Input split bytes=131
  Combine input records=62
  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)=402
  CPU time spent (ms)=4990
  Physical memory (bytes) snapshot=292900864
  Virtual memory (bytes) snapshot=4121735168
  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=133
File Output Format Counters
  Bytes Written=20
The mean is: 3.2903225806451615
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost mapreduce]$
```

Step3:

Display the contents of output file

```
Applications Places System Sun May 20, 1:46 PM Acadgild
acadgild@localhost:~/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce
File Edit View Search Terminal Help
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost mapreduce]$ hadoop fs -ls Assignment3/Task1/meanout
18/05/20 13:45:30 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 2018-05-20 11:57 Assignment3/Task1/meanout/_SUCCESS
-rw-r--r-- 1 acadgild supergroup 20 2018-05-20 11:57 Assignment3/Task1/meanout/part-r-00000
[acadgild@localhost mapreduce]$ hadoop fs -cat Assignment3/Task1/meanout/part-r-00000
18/05/20 13:45:55 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
count 31
length 102
[acadgild@localhost mapreduce]$
```

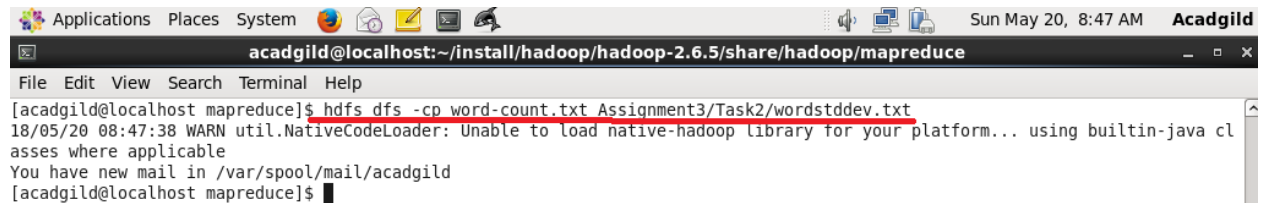
3. Word Standard Deviation

Step1:

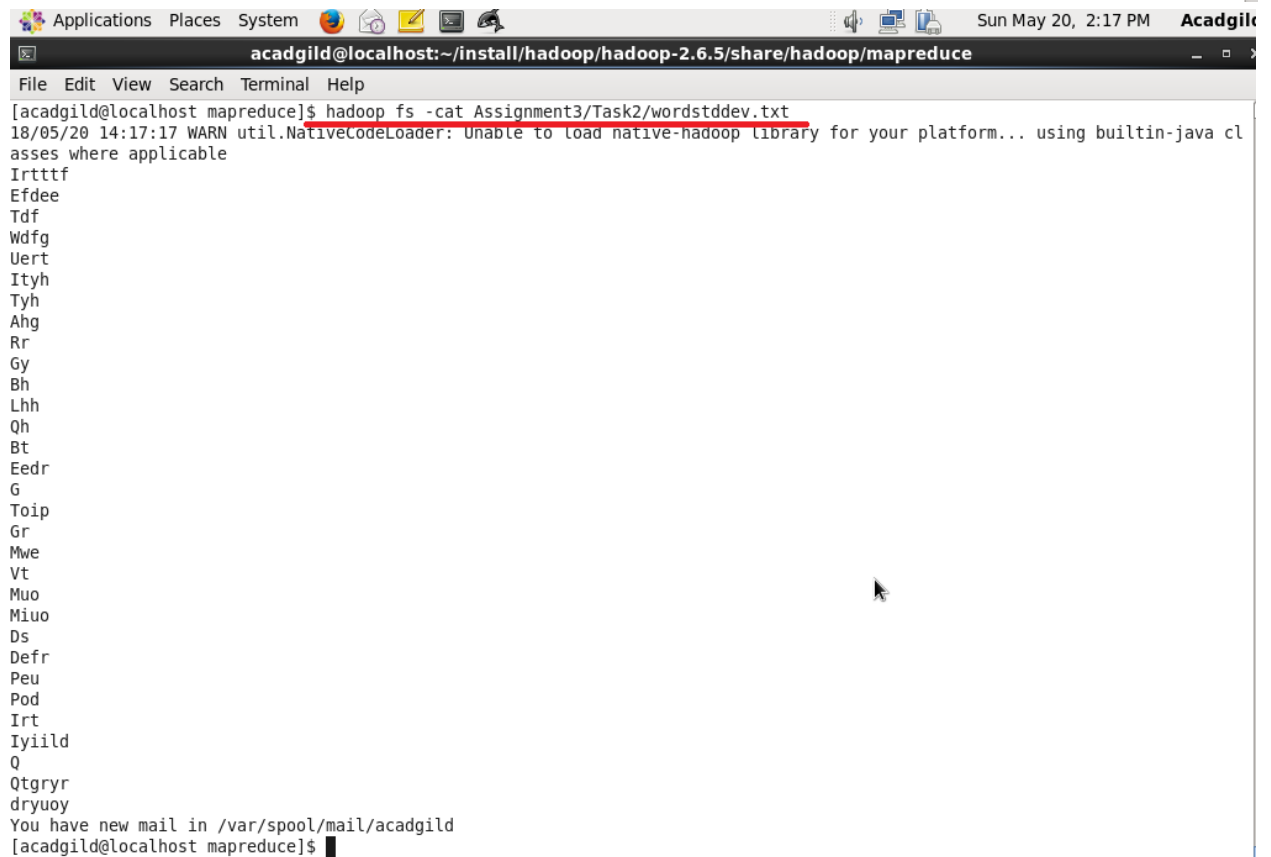
Copy the file wordcount.txt to wordstddev.txt

Command:

Hdfs dfs -cp wordcount.txt Assignment3/Task2/wordstddev.txt



```
Applications Places System 18/05/20, 8:47 AM Acadgild
acadgild@localhost: ~/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce
File Edit View Search Terminal Help
[acadgild@localhost mapreduce]$ hdfs dfs -cp word-count.txt Assignment3/Task2/wordstddev.txt
18/05/20 08:47:38 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost mapreduce]$
```



```
Applications Places System 18/05/20, 2:17 PM Acadgild
acadgild@localhost: ~/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce
File Edit View Search Terminal Help
[acadgild@localhost mapreduce]$ hadoop fs -cat Assignment3/Task2/wordstddev.txt
18/05/20 14:17:17 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
Irtttf
Efdee
Tdf
Wdfg
Uert
Ityh
Tyh
Ahg
Rr
Gy
Bh
Lhh
Qh
Bt
Eedr
G
Toip
Gr
Mwe
Vt
Muio
Miuo
Ds
Defr
Peu
Pod
Irt
Iyiild
Q
Qtgryr
dryuoy
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost mapreduce]$
```

Step2:

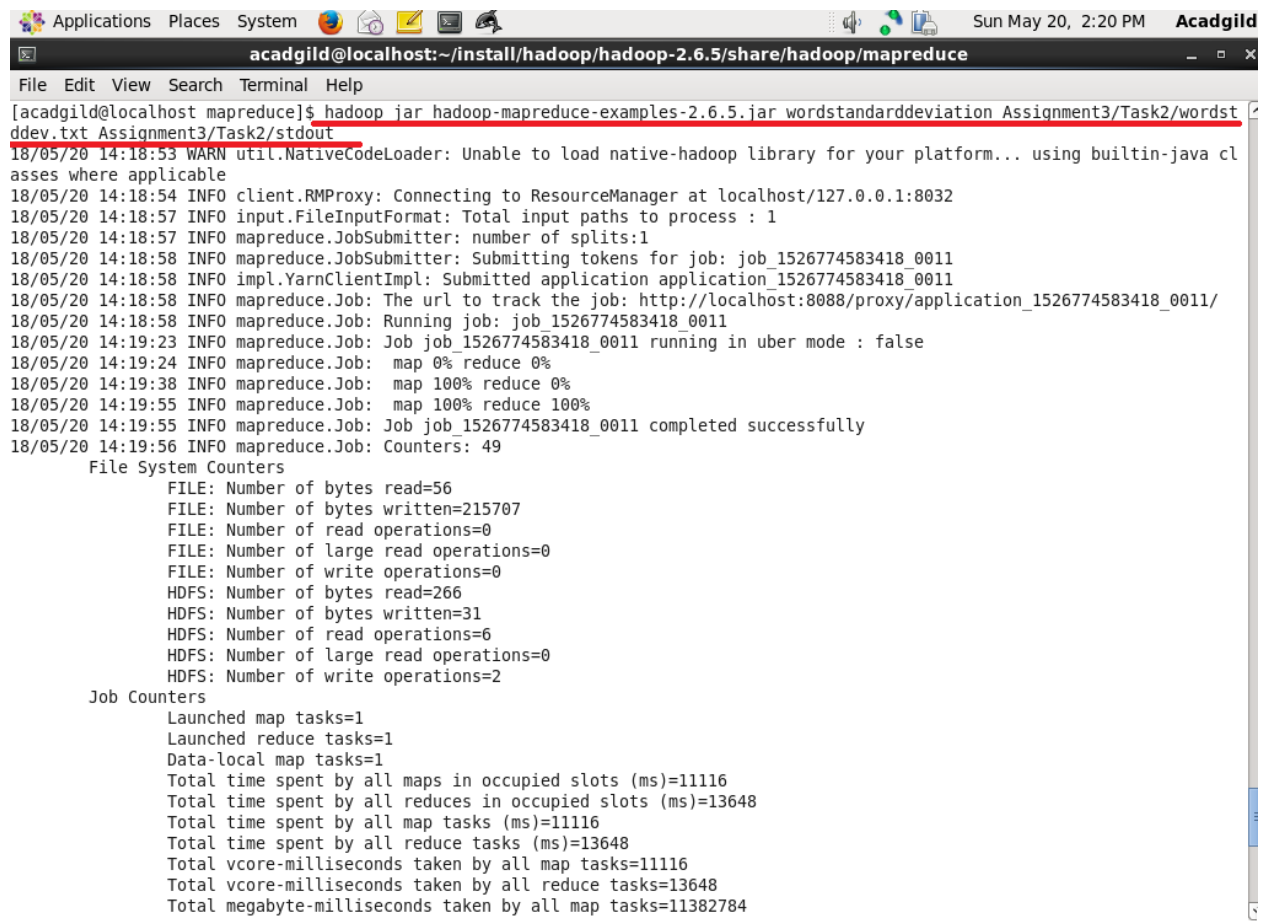
Execute the jar file for word standard deviation.

Syntax:

Hadoop jar <jarfilename> <mainclass name> <inputfile path>
<outputFilePath>

Command:

Hadoop jar Hadoop-mapreduce-examples-2.5.6.jar wordstandarddeviation
Assignment3/Task2/wordstddev.txt Assignment3/Task2/stdout



```
acadgild@localhost:~/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce
File Edit View Search Terminal Help
[acadgild@localhost mapreduce]$ hadoop jar hadoop-mapreduce-examples-2.6.5.jar wordstandarddeviation Assignment3/Task2/wordstd
ddev.txt Assignment3/Task2/stdout
18/05/20 14:18:53 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
18/05/20 14:18:54 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/05/20 14:18:57 INFO input.FileInputFormat: Total input paths to process : 1
18/05/20 14:18:57 INFO mapreduce.JobSubmitter: number of splits:1
18/05/20 14:18:58 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1526774583418_0011
18/05/20 14:18:58 INFO impl.YarnClientImpl: Submitted application application_1526774583418_0011
18/05/20 14:18:58 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1526774583418_0011/
18/05/20 14:18:58 INFO mapreduce.Job: Running job: job_1526774583418_0011
18/05/20 14:19:23 INFO mapreduce.Job: Job job_1526774583418_0011 running in uber mode : false
18/05/20 14:19:24 INFO mapreduce.Job: map 0% reduce 0%
18/05/20 14:19:38 INFO mapreduce.Job: map 100% reduce 0%
18/05/20 14:19:55 INFO mapreduce.Job: map 100% reduce 100%
18/05/20 14:19:55 INFO mapreduce.Job: Job job_1526774583418_0011 completed successfully
18/05/20 14:19:56 INFO mapreduce.Job: Counters: 49
  File System Counters
    FILE: Number of bytes read=56
    FILE: Number of bytes written=215707
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=266
    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)=11116
    Total time spent by all reduces in occupied slots (ms)=13648
    Total time spent by all map tasks (ms)=11116
    Total time spent by all reduce tasks (ms)=13648
    Total vcore-milliseconds taken by all map tasks=11116
    Total vcore-milliseconds taken by all reduce tasks=13648
    Total megabyte-milliseconds taken by all map tasks=11382784
```



```
Applications Places System Sun May 20, 2:20 PM Acadgild
acadgild@localhost:~/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce
File Edit View Search Terminal Help
Total vcore-milliseconds taken by all map tasks=11116
Total vcore-milliseconds taken by all reduce tasks=13648
Total megabyte-milliseconds taken by all map tasks=11382784
Total megabyte-milliseconds taken by all reduce tasks=13975552
Map-Reduce Framework
  Map input records=31
  Map output records=93
  Map output bytes=1364
  Map output materialized bytes=56
  Input split bytes=133
  Combine input records=93
  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)=211
  CPU time spent (ms)=2430
  Physical memory (bytes) snapshot=293580800
  Virtual memory (bytes) snapshot=4118228992
  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=133
File Output Format Counters
  Bytes Written=31
The standard deviation is: 1.3956972679215436
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost mapreduce]$
```

Step3:

Display the contents of output file

```
Applications Places System Sun May 20, 2:21 PM Acadgild
acadgild@localhost:~/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce
File Edit View Search Terminal Help
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost mapreduce]$ hadoop fs -ls Assignment3/Task2/stdout
18/05/20 14:21:22 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 2018-05-20 14:19 Assignment3/Task2/stdout/ SUCCESS
-rw-r--r-- 1 acadgild supergroup 31 2018-05-20 14:19 Assignment3/Task2/stdout/part-r-00000
[acadgild@localhost mapreduce]$ hadoop fs -cat Assignment3/Task2/stdout/part-r-00000
18/05/20 14:21:30 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
count 31
length 102
square 396
[acadgild@localhost mapreduce]$
```