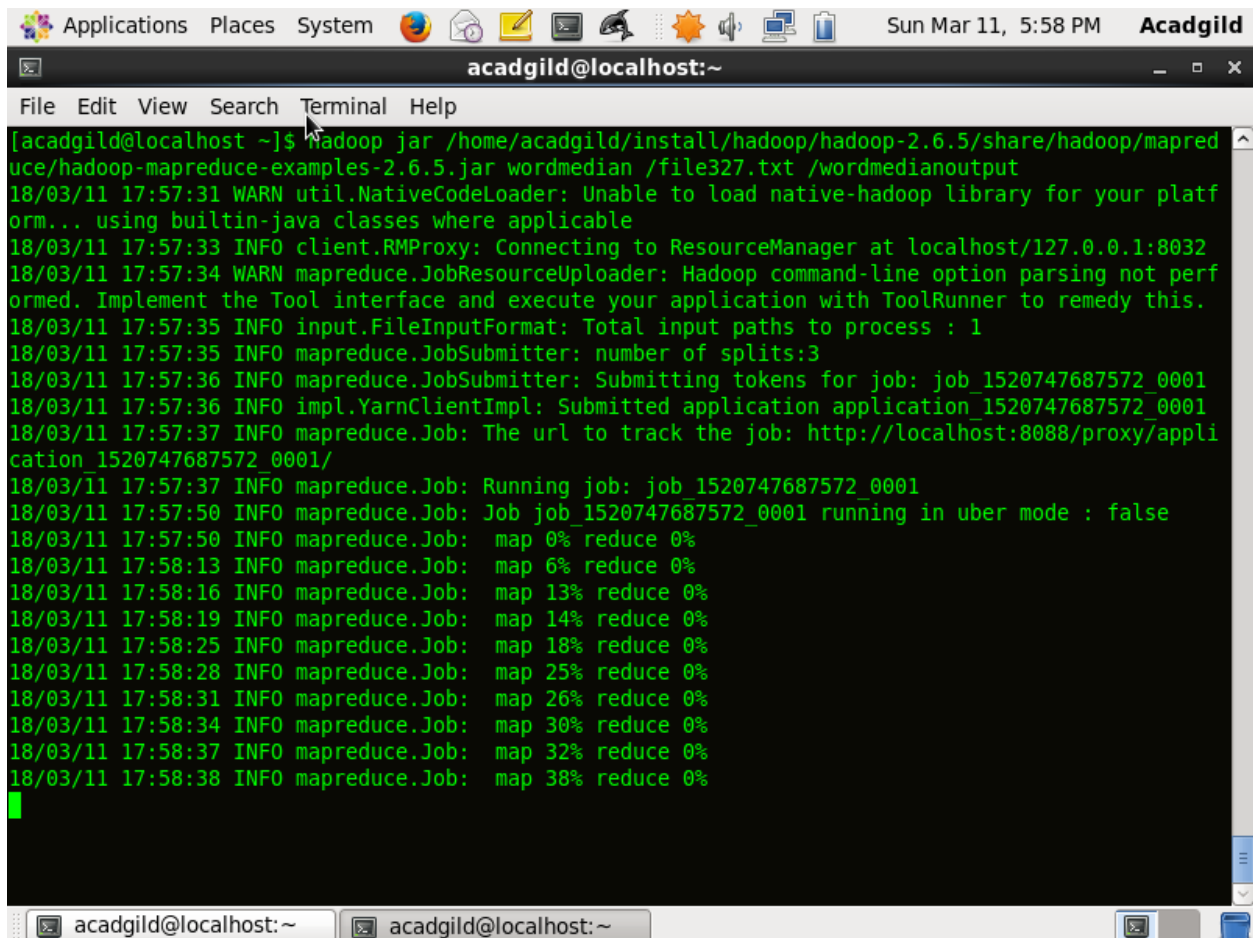


To run the wordmedian map reduce program enter the command **hadoop jar /home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce/hadoop-mapreduce-examples-2.6.5.jar wordmedian /file327.txt /wordmedianoutput**

Here, file327.txt is the text file whose median will be found out and the output will be stored in the directory **/wordmedianoutput**

Alternatively the path of the jar file could have been specified by

\$HADOOP_HOME/share/hadoop/mapreduce/hadoop-mapreduce-examples-2.6.5.jar

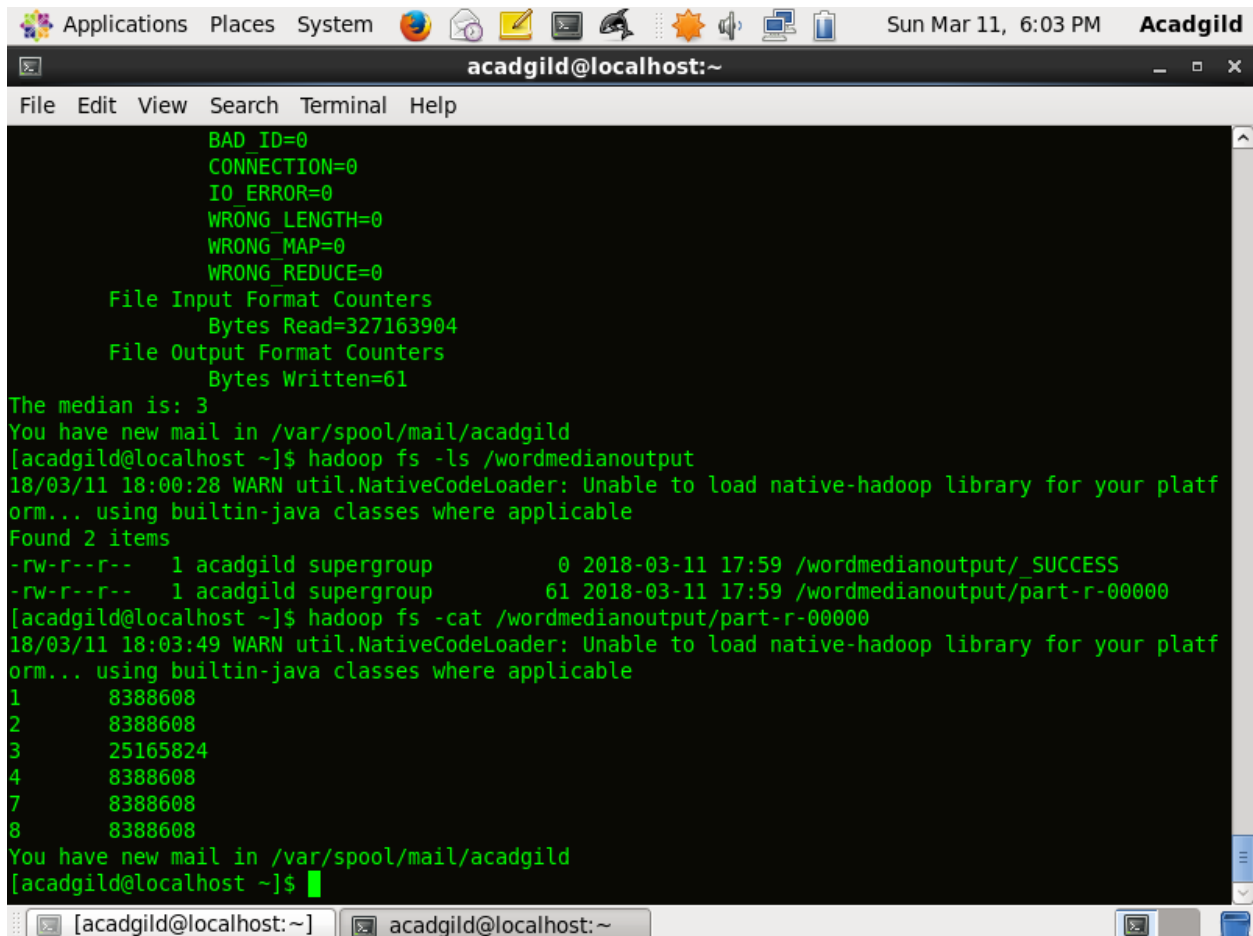


The screenshot shows a terminal window titled "acadgild@localhost:~" with a menu bar (File, Edit, View, Search, Terminal, Help) and a toolbar. The terminal output displays the execution of the Hadoop wordmedian job. It starts with the command: `[acadgild@localhost ~]$ hadoop jar /home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce/hadoop-mapreduce-examples-2.6.5.jar wordmedian /file327.txt /wordmedianoutput`. The output includes several log messages: a warning about the native Hadoop library, information about connecting to the ResourceManager, a warning about command-line option parsing, and information about the job submission and progress. The job is identified as `job_1520747687572_0001`. The progress shows the map phase reaching 38% completion, while the reduce phase remains at 0%.

```
[acadgild@localhost ~]$ hadoop jar /home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce/hadoop-mapreduce-examples-2.6.5.jar wordmedian /file327.txt /wordmedianoutput
18/03/11 17:57:31 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/03/11 17:57:33 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/03/11 17:57:34 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
18/03/11 17:57:35 INFO input.FileInputFormat: Total input paths to process : 1
18/03/11 17:57:35 INFO mapreduce.JobSubmitter: number of splits:3
18/03/11 17:57:36 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1520747687572_0001
18/03/11 17:57:36 INFO impl.YarnClientImpl: Submitted application application_1520747687572_0001
18/03/11 17:57:37 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1520747687572_0001/
18/03/11 17:57:37 INFO mapreduce.Job: Running job: job_1520747687572_0001
18/03/11 17:57:50 INFO mapreduce.Job: Job job_1520747687572_0001 running in uber mode : false
18/03/11 17:57:50 INFO mapreduce.Job: map 0% reduce 0%
18/03/11 17:58:13 INFO mapreduce.Job: map 6% reduce 0%
18/03/11 17:58:16 INFO mapreduce.Job: map 13% reduce 0%
18/03/11 17:58:19 INFO mapreduce.Job: map 14% reduce 0%
18/03/11 17:58:25 INFO mapreduce.Job: map 18% reduce 0%
18/03/11 17:58:28 INFO mapreduce.Job: map 25% reduce 0%
18/03/11 17:58:31 INFO mapreduce.Job: map 26% reduce 0%
18/03/11 17:58:34 INFO mapreduce.Job: map 30% reduce 0%
18/03/11 17:58:37 INFO mapreduce.Job: map 32% reduce 0%
18/03/11 17:58:38 INFO mapreduce.Job: map 38% reduce 0%
```

Use command **hadoop fs -ls /wordmedianoutput** and **hadoop fs -cat /wordmedianoutput/part-r-00000** to see your reduced data which is used to compute the median value.

The median is **3** as seen in the screenshot

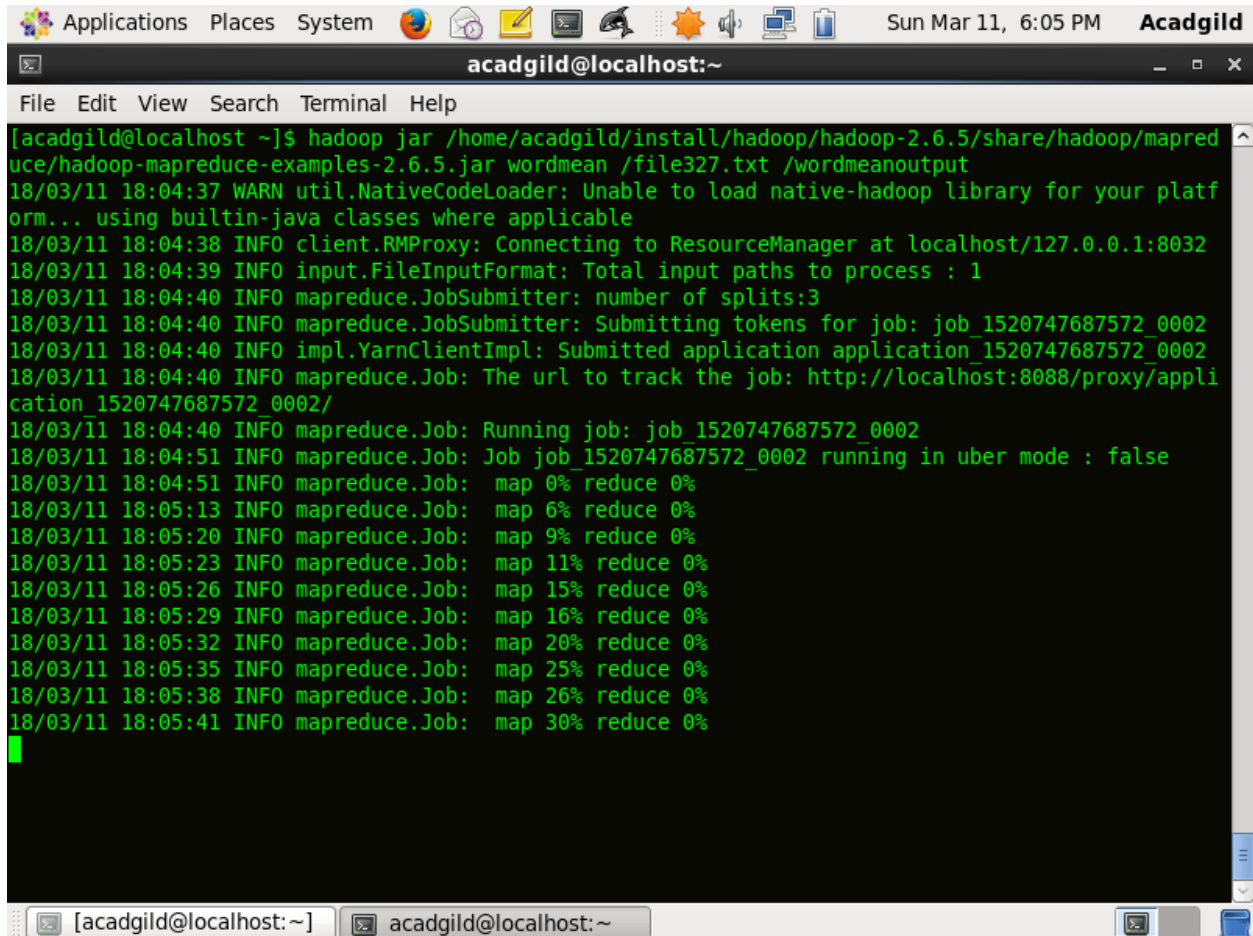


The screenshot shows a terminal window titled 'acadgild@localhost:~' with a menu bar (File, Edit, View, Search, Terminal, Help) and a system bar at the top (Applications, Places, System, icons, Sun Mar 11, 6:03 PM, Acadgild). The terminal output is as follows:

```
BAD_ID=0
CONNECTION=0
IO_ERROR=0
WRONG_LENGTH=0
WRONG_MAP=0
WRONG_REDUCE=0
File Input Format Counters
  Bytes Read=327163904
File Output Format Counters
  Bytes Written=61
The median is: 3
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$ hadoop fs -ls /wordmedianoutput
18/03/11 18:00:28 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platf
orm... using builtin-java classes where applicable
Found 2 items
-rw-r--r--  1 acadgild supergroup          0 2018-03-11 17:59 /wordmedianoutput/_SUCCESS
-rw-r--r--  1 acadgild supergroup        61 2018-03-11 17:59 /wordmedianoutput/part-r-00000
[acadgild@localhost ~]$ hadoop fs -cat /wordmedianoutput/part-r-00000
18/03/11 18:03:49 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platf
orm... using builtin-java classes where applicable
1      8388608
2      8388608
3     25165824
4      8388608
7      8388608
8      8388608
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$
```

To run the wordmedian map reduce program enter the command **hadoop jar /home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce/hadoop-mapreduce-examples-2.6.5.jar wordmean /file327.txt /wordmeanoutput**

Here, file327.txt is the text file whose mean will be found out and the output will be stored in the directory /wordmeanoutput



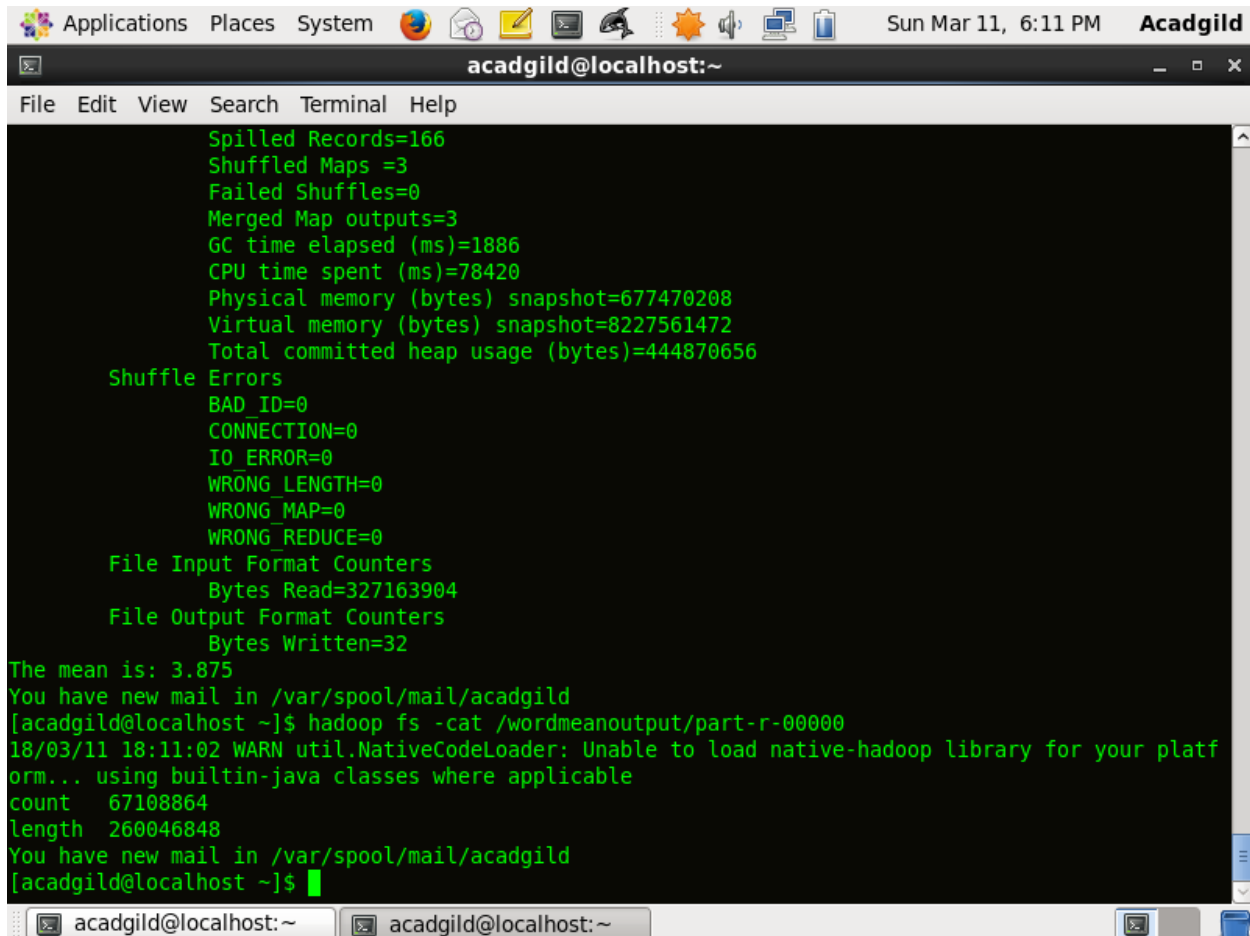
The screenshot shows a terminal window titled 'acadgild@localhost:~' with a menu bar (File, Edit, View, Search, Terminal, Help). The terminal displays the command to run the Hadoop wordmean job and its output. The output shows the job submission and progress updates.

```
[acadgild@localhost ~]$ hadoop jar /home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce/hadoop-mapreduce-examples-2.6.5.jar wordmean /file327.txt /wordmeanoutput
18/03/11 18:04:37 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/03/11 18:04:38 INFO client.RMProxy: Connecting to Resource Manager at localhost/127.0.0.1:8032
18/03/11 18:04:39 INFO input.FileInputFormat: Total input paths to process : 1
18/03/11 18:04:40 INFO mapreduce.JobSubmitter: number of splits:3
18/03/11 18:04:40 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1520747687572_0002
18/03/11 18:04:40 INFO impl.YarnClientImpl: Submitted application application_1520747687572_0002
18/03/11 18:04:40 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1520747687572_0002/
18/03/11 18:04:40 INFO mapreduce.Job: Running job: job_1520747687572_0002
18/03/11 18:04:51 INFO mapreduce.Job: Job job_1520747687572_0002 running in uber mode : false
18/03/11 18:04:51 INFO mapreduce.Job: map 0% reduce 0%
18/03/11 18:05:13 INFO mapreduce.Job: map 6% reduce 0%
18/03/11 18:05:20 INFO mapreduce.Job: map 9% reduce 0%
18/03/11 18:05:23 INFO mapreduce.Job: map 11% reduce 0%
18/03/11 18:05:26 INFO mapreduce.Job: map 15% reduce 0%
18/03/11 18:05:29 INFO mapreduce.Job: map 16% reduce 0%
18/03/11 18:05:32 INFO mapreduce.Job: map 20% reduce 0%
18/03/11 18:05:35 INFO mapreduce.Job: map 25% reduce 0%
18/03/11 18:05:38 INFO mapreduce.Job: map 26% reduce 0%
18/03/11 18:05:41 INFO mapreduce.Job: map 30% reduce 0%
```

Output destination is /wordmeanoutput so run the command **hadoop fs -ls /wordmeanoutput** to explore the output directory

You can see output of the reducer in the part-r-00000 file so run the command **hadoop fs -cat /wordmeanoutput/part-r-00000**

The mean is **3.875** which is equal to value of length divided by count as given by the reducer



The screenshot shows a terminal window titled 'acadgild@localhost:~' with a menu bar (File, Edit, View, Search, Terminal, Help) and a system bar at the top (Applications, Places, System, icons, Sun Mar 11, 6:11 PM, Acadgild). The terminal output displays Hadoop job statistics in green text, followed by file input/output counters, and the results of a 'cat' command on a HDFS file. The results show a count of 67108864 and a length of 260046848. A warning message about the native Hadoop library is also visible.

```
Spilled Records=166
Shuffled Maps =3
Failed Shuffles=0
Merged Map outputs=3
GC time elapsed (ms)=1886
CPU time spent (ms)=78420
Physical memory (bytes) snapshot=677470208
Virtual memory (bytes) snapshot=8227561472
Total committed heap usage (bytes)=444870656

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=327163904
File Output Format Counters
  Bytes Written=32

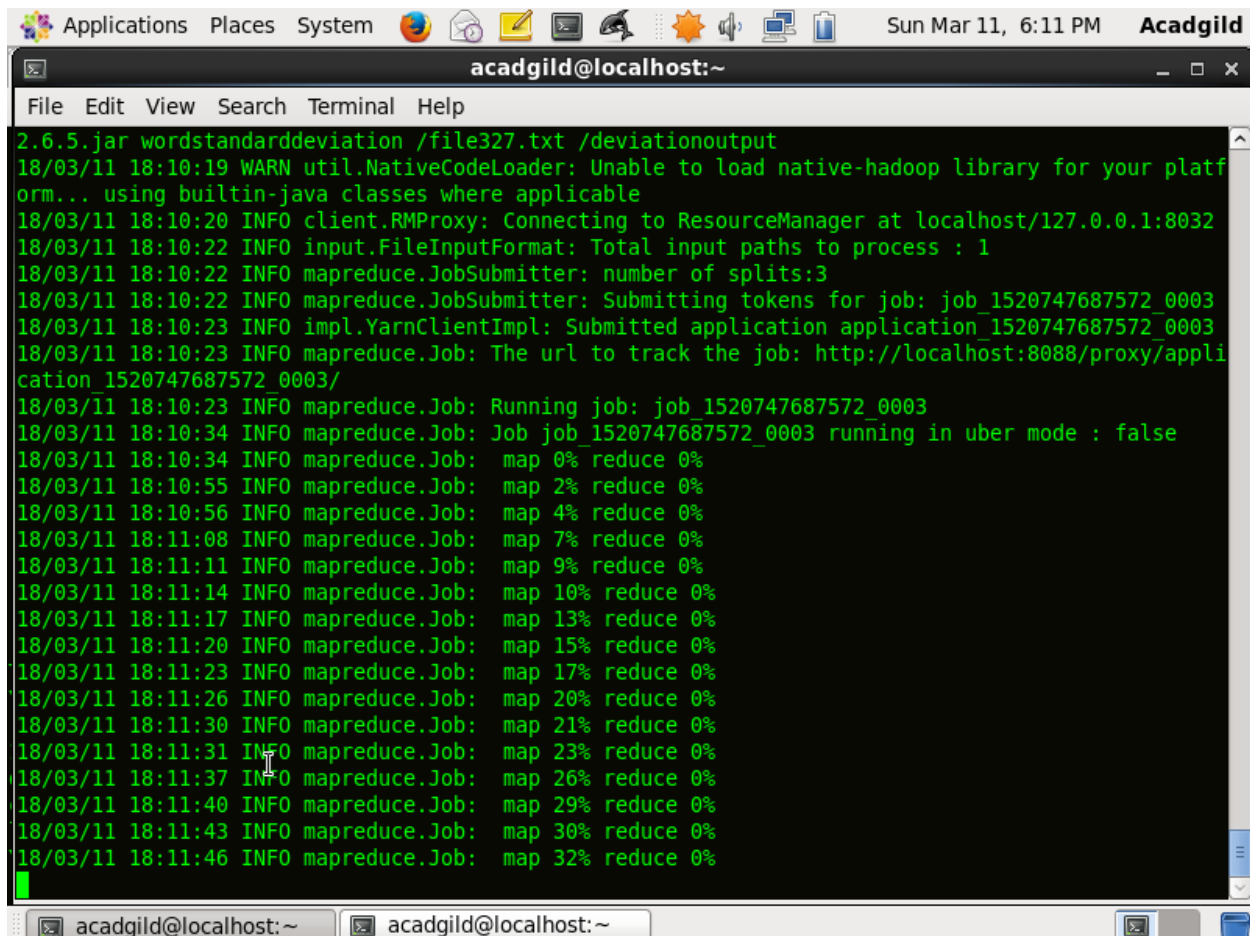
The mean is: 3.875
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$ hadoop fs -cat /wordmeanoutput/part-r-00000
18/03/11 18:11:02 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platf
orm... using builtin-java classes where applicable
count    67108864
length   260046848
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$
```

To run the standard deviation map reduce program enter the command **hadoop jar /home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce/hadoop-mapreduce-examples-2.6.5.jar wordstandarddeviation /file327.txt /deviationoutput**

Here, file327.txt is the text file whose standard deviation words will be found out and the output will be stored in the directory /deviationoutput

Alternatively the path of the jar file could have been specified by

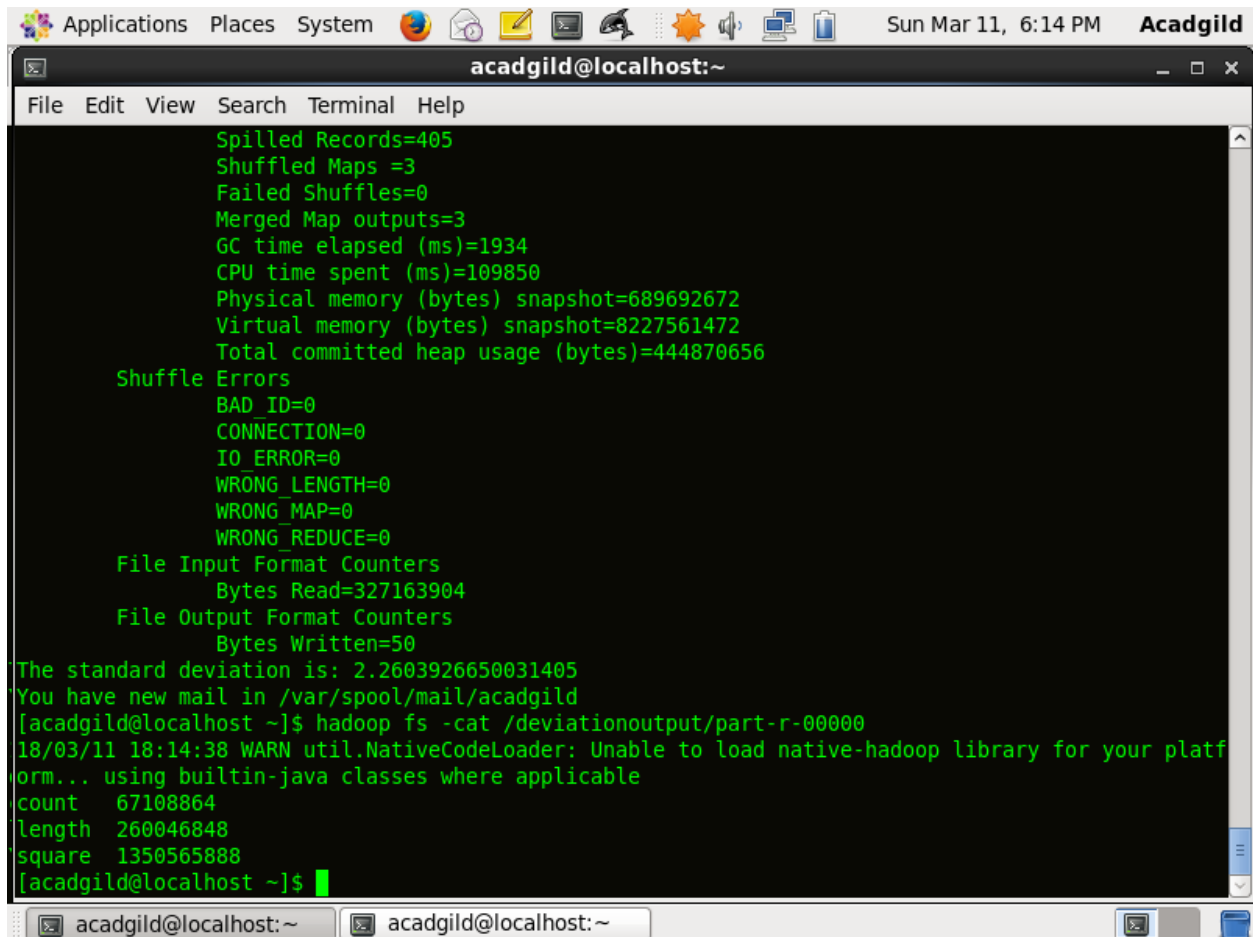
\$HADOOP_HOME/share/hadoop/mapreduce/hadoop-mapreduce-examples-2.6.5.jar



```
acacdgild@localhost:~  
File Edit View Search Terminal Help  
2.6.5.jar wordstandarddeviation /file327.txt /deviationoutput  
18/03/11 18:10:19 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platf  
orm... using builtin-java classes where applicable  
18/03/11 18:10:20 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032  
18/03/11 18:10:22 INFO input.FileInputFormat: Total input paths to process : 1  
18/03/11 18:10:22 INFO mapreduce.JobSubmitter: number of splits:3  
18/03/11 18:10:22 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1520747687572_0003  
18/03/11 18:10:23 INFO impl.YarnClientImpl: Submitted application application_1520747687572_0003  
18/03/11 18:10:23 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/appli  
cation_1520747687572_0003/  
18/03/11 18:10:23 INFO mapreduce.Job: Running job: job_1520747687572_0003  
18/03/11 18:10:34 INFO mapreduce.Job: Job job_1520747687572_0003 running in uber mode : false  
18/03/11 18:10:34 INFO mapreduce.Job: map 0% reduce 0%  
18/03/11 18:10:55 INFO mapreduce.Job: map 2% reduce 0%  
18/03/11 18:10:56 INFO mapreduce.Job: map 4% reduce 0%  
18/03/11 18:11:08 INFO mapreduce.Job: map 7% reduce 0%  
18/03/11 18:11:11 INFO mapreduce.Job: map 9% reduce 0%  
18/03/11 18:11:14 INFO mapreduce.Job: map 10% reduce 0%  
18/03/11 18:11:17 INFO mapreduce.Job: map 13% reduce 0%  
18/03/11 18:11:20 INFO mapreduce.Job: map 15% reduce 0%  
18/03/11 18:11:23 INFO mapreduce.Job: map 17% reduce 0%  
18/03/11 18:11:26 INFO mapreduce.Job: map 20% reduce 0%  
18/03/11 18:11:30 INFO mapreduce.Job: map 21% reduce 0%  
18/03/11 18:11:31 INFO mapreduce.Job: map 23% reduce 0%  
18/03/11 18:11:37 INFO mapreduce.Job: map 26% reduce 0%  
18/03/11 18:11:40 INFO mapreduce.Job: map 29% reduce 0%  
18/03/11 18:11:43 INFO mapreduce.Job: map 30% reduce 0%  
18/03/11 18:11:46 INFO mapreduce.Job: map 32% reduce 0%
```

Screenshot of output:

Standard deviation is obtained as 2.603



The screenshot shows a Linux desktop environment with a terminal window titled 'acadgild@localhost:~'. The terminal displays the output of a Hadoop job. The output includes various statistics such as 'Spilled Records=405', 'Shuffled Maps =3', 'Failed Shuffles=0', 'Merged Map outputs=3', 'GC time elapsed (ms)=1934', 'CPU time spent (ms)=109850', 'Physical memory (bytes) snapshot=689692672', 'Virtual memory (bytes) snapshot=8227561472', and 'Total committed heap usage (bytes)=444870656'. It also lists 'Shuffle Errors' which are all zero, and 'File Input Format Counters' showing 'Bytes Read=327163904' and 'File Output Format Counters' showing 'Bytes Written=50'. The standard deviation is calculated as 2.2603926650031405. A mail notification is shown: 'You have new mail in /var/spool/mail/acadgild'. The user then runs the command 'hadoop fs -cat /deviationoutput/part-r-00000', which outputs a large number of lines of data, including 'count 67108864', 'length 260046848', and 'square 1350565888'. The terminal window has a menu bar with 'File', 'Edit', 'View', 'Search', 'Terminal', and 'Help'. The desktop background is black, and the taskbar at the bottom shows the terminal window and a few icons.

```
Applications Places System Sun Mar 11, 6:14 PM Acadgild
acadgild@localhost:~
File Edit View Search Terminal Help
Spilled Records=405
Shuffled Maps =3
Failed Shuffles=0
Merged Map outputs=3
GC time elapsed (ms)=1934
CPU time spent (ms)=109850
Physical memory (bytes) snapshot=689692672
Virtual memory (bytes) snapshot=8227561472
Total committed heap usage (bytes)=444870656
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=327163904
File Output Format Counters
Bytes Written=50
The standard deviation is: 2.2603926650031405
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$ hadoop fs -cat /deviationoutput/part-r-00000
18/03/11 18:14:38 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platf
orm... using builtin-java classes where applicable
count 67108864
length 260046848
square 1350565888
[acadgild@localhost ~]$
```