

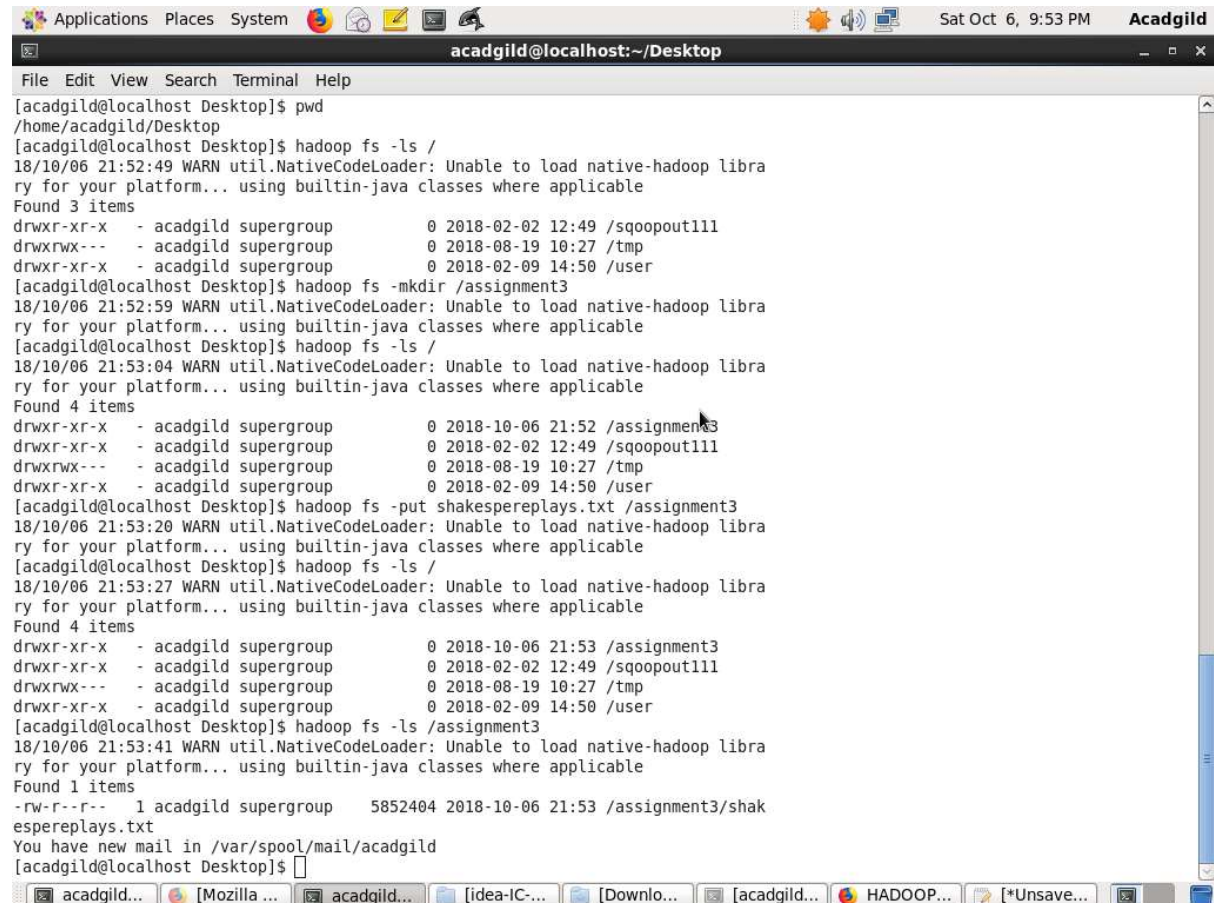
Step 1:

Downloading a text file from Internet which has all Shakespeare Plays.

URL - <http://www.gutenberg.org/files/100/100-0.txt>

Step 2:

Since the text file is not in the HDFS, copying this file using the commands as shown in the screenshot:



The screenshot shows a terminal window titled 'acadgild@localhost:~/Desktop'. The user is logged in as 'acadgild'. The terminal displays the following commands and output:

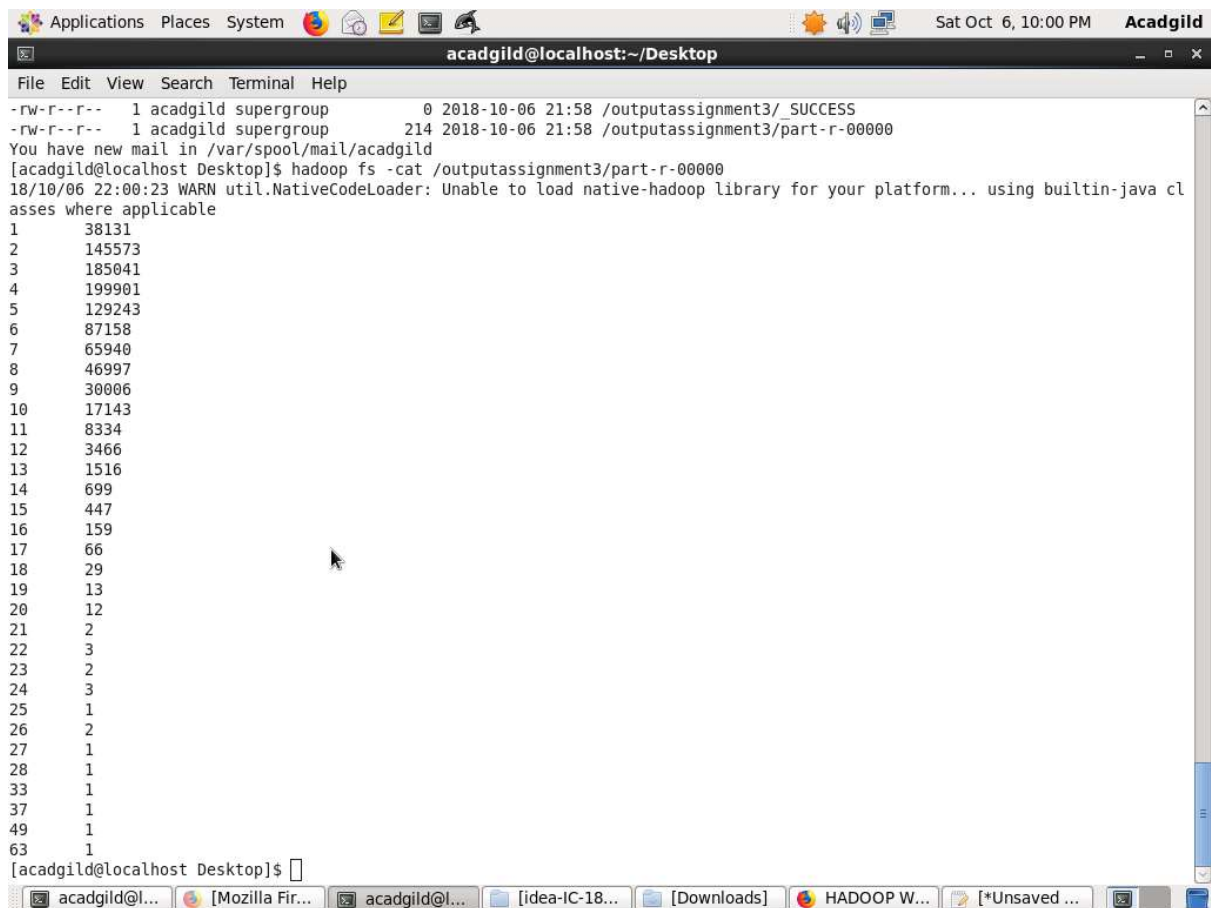
```
[acadgild@localhost Desktop]$ pwd
/home/acadgild/Desktop
[acadgild@localhost Desktop]$ hadoop fs -ls /
18/10/06 21:52:49 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 3 items
drwxr-xr-x - acadgild supergroup          0 2018-02-02 12:49 /sqoopout111
drwxrwx--- - acadgild supergroup          0 2018-08-19 10:27 /tmp
drwxr-xr-x - acadgild supergroup          0 2018-02-09 14:50 /user
[acadgild@localhost Desktop]$ hadoop fs -mkdir /assignment3
18/10/06 21:52:59 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
[acadgild@localhost Desktop]$ hadoop fs -ls /
18/10/06 21:53:04 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 4 items
drwxr-xr-x - acadgild supergroup          0 2018-10-06 21:52 /assignment3
drwxr-xr-x - acadgild supergroup          0 2018-02-02 12:49 /sqoopout111
drwxrwx--- - acadgild supergroup          0 2018-08-19 10:27 /tmp
drwxr-xr-x - acadgild supergroup          0 2018-02-09 14:50 /user
[acadgild@localhost Desktop]$ hadoop fs -put shakespeareplays.txt /assignment3
18/10/06 21:53:20 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
[acadgild@localhost Desktop]$ hadoop fs -ls /
18/10/06 21:53:27 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 4 items
drwxr-xr-x - acadgild supergroup          0 2018-10-06 21:53 /assignment3
drwxr-xr-x - acadgild supergroup          0 2018-02-02 12:49 /sqoopout111
drwxrwx--- - acadgild supergroup          0 2018-08-19 10:27 /tmp
drwxr-xr-x - acadgild supergroup          0 2018-02-09 14:50 /user
[acadgild@localhost Desktop]$ hadoop fs -ls /assignment3
18/10/06 21:53:41 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 1 items
-rw-r--r-- 1 acadgild supergroup    5852404 2018-10-06 21:53 /assignment3/shakespeareplays.txt
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost Desktop]$
```

Step 3:

Executing the WordMedian program as shown in the following screenshots:

```
Applications Places System Sat Oct 6, 9:53 PM Acadgild
acadgild@localhost:~/Desktop
File Edit View Search Terminal Help
[acadgild@localhost Desktop]$ pwd
/home/acadgild/Desktop
[acadgild@localhost Desktop]$ hadoop fs -ls /
18/10/06 21:52:49 WARN util.NativeCodeLoader: Unable to load native-hadoop libra
ry for your platform... using builtin-java classes where applicable
Found 3 items
drwxr-xr-x - acadgild supergroup 0 2018-02-02 12:49 /sqoopout111
drwxrwx-- - acadgild supergroup 0 2018-08-19 10:27 /tmp
drwxr-xr-x - acadgild supergroup 0 2018-02-09 14:50 /user
[acadgild@localhost Desktop]$ hadoop fs -mkdir /assignment3
18/10/06 21:52:59 WARN util.NativeCodeLoader: Unable to load native-hadoop libra
ry for your platform... using builtin-java classes where applicable
[acadgild@localhost Desktop]$ hadoop fs -ls /
18/10/06 21:53:04 WARN util.NativeCodeLoader: Unable to load native-hadoop libra
ry for your platform... using builtin-java classes where applicable
Found 4 items
drwxr-xr-x - acadgild supergroup 0 2018-10-06 21:52 /assignment3
drwxr-xr-x - acadgild supergroup 0 2018-02-02 12:49 /sqoopout111
drwxrwx-- - acadgild supergroup 0 2018-08-19 10:27 /tmp
drwxr-xr-x - acadgild supergroup 0 2018-02-09 14:50 /user
[acadgild@localhost Desktop]$ hadoop fs -put shakespeareplays.txt /assignment3
18/10/06 21:53:20 WARN util.NativeCodeLoader: Unable to load native-hadoop libra
ry for your platform... using builtin-java classes where applicable
[acadgild@localhost Desktop]$ hadoop fs -ls /
18/10/06 21:53:27 WARN util.NativeCodeLoader: Unable to load native-hadoop libra
ry for your platform... using builtin-java classes where applicable
Found 4 items
drwxr-xr-x - acadgild supergroup 0 2018-10-06 21:53 /assignment3
drwxr-xr-x - acadgild supergroup 0 2018-02-02 12:49 /sqoopout111
drwxrwx-- - acadgild supergroup 0 2018-08-19 10:27 /tmp
drwxr-xr-x - acadgild supergroup 0 2018-02-09 14:50 /user
[acadgild@localhost Desktop]$ hadoop fs -ls /assignment3
18/10/06 21:53:41 WARN util.NativeCodeLoader: Unable to load native-hadoop libra
ry for your platform... using builtin-java classes where applicable
Found 1 items
-rw-r--r-- 1 acadgild supergroup 5852404 2018-10-06 21:53 /assignment3/shak
espereplays.txt
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost Desktop]$
```

```
Applications Places System Sat Oct 6, 9:58 PM Acadgild
acadgild@localhost:~/Desktop
File Edit View Search Terminal Help
Total vcore-milliseconds taken by all map tasks=2852
Total vcore-milliseconds taken by all reduce tasks=2547
Total megabyte-milliseconds taken by all map tasks=2920448
Total megabyte-milliseconds taken by all reduce tasks=2608128
Map-Reduce Framework
Map input records=149689
Map output records=959893
Map output bytes=7679144
Map output materialized bytes=326
Input split bytes=118
Combine input records=959893
Combine output records=32
Reduce input groups=32
Reduce shuffle bytes=326
Reduce input records=32
Reduce output records=32
Spilled Records=64
Shuffled Maps =1
Failed Shuffles=0
Merged Map outputs=1
GC time elapsed (ms)=89
CPU time spent (ms)=2550
Physical memory (bytes) snapshot=444801024
Virtual memory (bytes) snapshot=4165038080
Total committed heap usage (bytes)=300417024
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=5852404
File Output Format Counters
Bytes Written=214
The median is: 4
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost Desktop]$
```



The screenshot shows a terminal window titled 'acadgild@localhost:~/Desktop'. The terminal output displays Hadoop file system information for a file named 'part-r-00000'. It shows permissions, owner (acadgild supergroup), size (214), and timestamp (2018-10-06 21:58). A warning message indicates that the native Hadoop library could not be loaded, so the built-in Java class loader is being used. Below this, a list of 63 blocks is shown, each with a block ID and its size in bytes.

```
-rw-r--r-- 1 acadgild supergroup 0 2018-10-06 21:58 /outputassignment3/_SUCCESS
-rw-r--r-- 1 acadgild supergroup 214 2018-10-06 21:58 /outputassignment3/part-r-00000
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost Desktop]$ hadoop fs -cat /outputassignment3/part-r-00000
18/10/06 22:00:23 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
1      38131
2      145573
3      185041
4      199901
5      129243
6       87158
7       65940
8       46997
9       30006
10     17143
11     8334
12     3466
13     1516
14      699
15     447
16     159
17      66
18     29
19     13
20     12
21      2
22      3
23      2
24      3
25      1
26      2
27      1
28      1
33      1
37      1
49      1
63      1
[acadgild@localhost Desktop]$
```

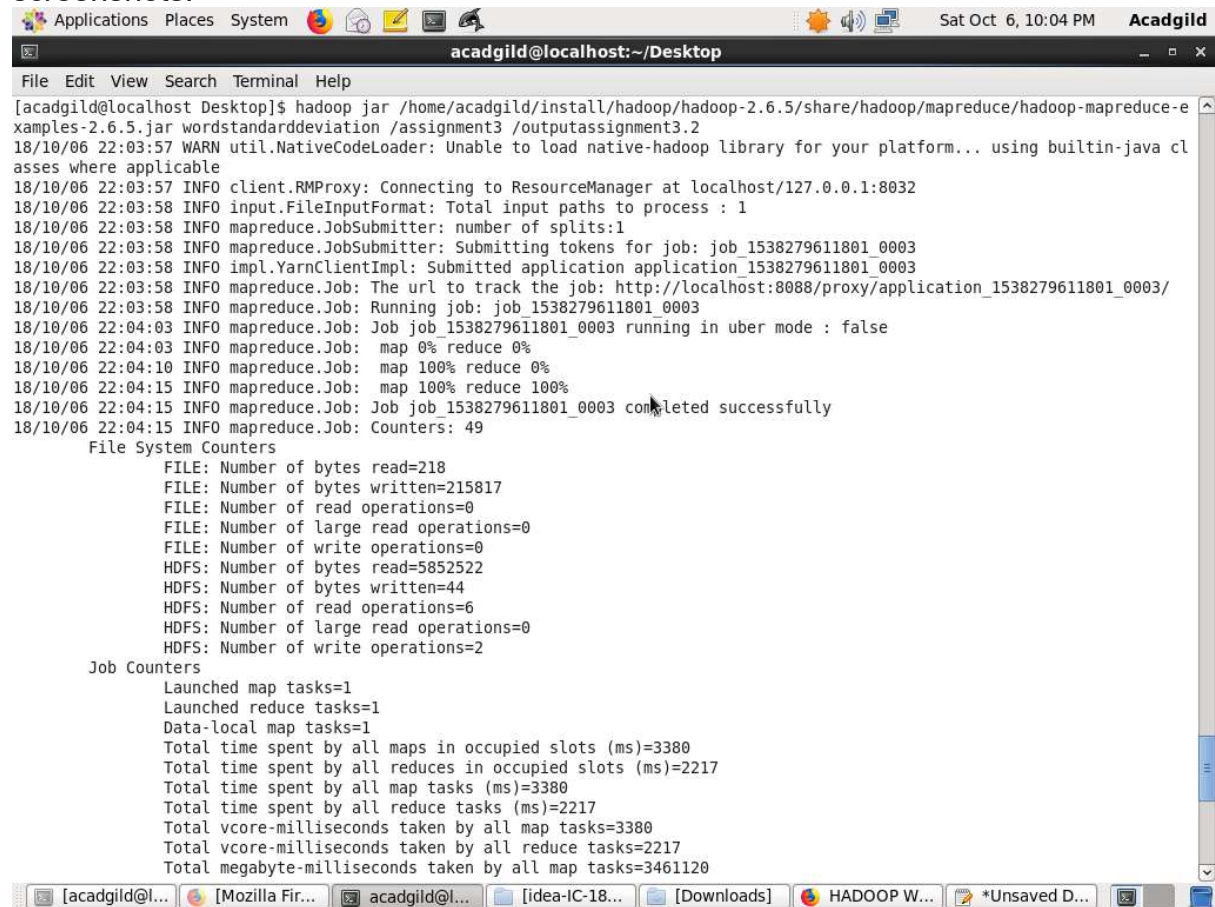
Step 4:
Executing the WordMean program as shown in the following screenshots:

```
Applications Places System Sat Oct 6, 10:01 PM Acadgild
acadgild@localhost:~/Desktop
File Edit View Search Terminal Help
[acadgild@localhost Desktop]$ hadoop jar /home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce/hadoop-mapreduce-examples-2.6.5.jar wordmean /assignment3/outputassignment3.1
18/10/06 22:01:01 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/10/06 22:01:02 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/10/06 22:01:03 INFO input.FileInputFormat: Total input paths to process : 1
18/10/06 22:01:03 INFO mapreduce.JobSubmitter: number of splits:1
18/10/06 22:01:03 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1538279611801_0002
18/10/06 22:01:03 INFO impl.YarnClientImpl: Submitted application application_1538279611801_0002
18/10/06 22:01:03 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1538279611801_0002/
18/10/06 22:01:03 INFO mapreduce.Job: Running job: job_1538279611801_0002
18/10/06 22:01:08 INFO mapreduce.Job: Job job_1538279611801_0002 running in uber mode : false
18/10/06 22:01:08 INFO mapreduce.Job: map 0% reduce 0%
18/10/06 22:01:13 INFO mapreduce.Job: map 100% reduce 0%
18/10/06 22:01:18 INFO mapreduce.Job: map 100% reduce 100%
18/10/06 22:01:19 INFO mapreduce.Job: Job job_1538279611801_0002 completed successfully
18/10/06 22:01:19 INFO mapreduce.Job: Counters: 49
  File System Counters
    FILE: Number of bytes read=39
    FILE: Number of bytes written=215411
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=5852522
    HDFS: Number of bytes written=28
    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)=3013
    Total time spent by all reduces in occupied slots (ms)=2183
    Total time spent by all map tasks (ms)=3013
    Total time spent by all reduce tasks (ms)=2183
    Total vcore-milliseconds taken by all map tasks=3013
    Total vcore-milliseconds taken by all reduce tasks=2183
    Total megabyte-milliseconds taken by all map tasks=3085312
```

```
Applications Places System Sat Oct 6, 10:02 PM Acadgild
acadgild@localhost:~/Desktop
File Edit View Search Terminal Help
  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)=105
  CPU time spent (ms)=2540
  Physical memory (bytes) snapshot=441245696
  Virtual memory (bytes) snapshot=4169424896
  Total committed heap usage (bytes)=293601280
  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=5852404
  File Output Format Counters
    Bytes Written=28
  The mean is: 4.486840720788671
  You have new mail in /var/spool/mail/acadgild
[acadgild@localhost Desktop]$ hadoop fs -ls /outputassignment3.1
18/10/06 22:01:48 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 2018-10-06 22:01 /outputassignment3.1/ SUCCESS
-rw-r--r-- 1 acadgild supergroup 28 2018-10-06 22:01 /outputassignment3.1/part-r-00000
[acadgild@localhost Desktop]$ hadoop fs -cat /outputassignment3.1/part-r-00000
18/10/06 22:02:01 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
count 959893
length 4306887
[acadgild@localhost Desktop]$
```


Step 5:

Executing the WordStandardDeviation program as shown in the following screenshots:



```
[acadgild@localhost Desktop]$ hadoop jar /home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce/hadoop-mapreduce-examples-2.6.5.jar wordstandarddeviation /assignment3 /outputassignment3.2
18/10/06 22:03:57 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/10/06 22:03:57 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/10/06 22:03:58 INFO input.FileInputFormat: Total input paths to process : 1
18/10/06 22:03:58 INFO mapreduce.JobSubmitter: number of splits:1
18/10/06 22:03:58 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1538279611801_0003
18/10/06 22:03:58 INFO impl.YarnClientImpl: Submitted application application_1538279611801_0003
18/10/06 22:03:58 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1538279611801_0003/
18/10/06 22:03:58 INFO mapreduce.Job: Running job: job_1538279611801_0003
18/10/06 22:04:03 INFO mapreduce.Job: Job job_1538279611801_0003 running in uber mode : false
18/10/06 22:04:03 INFO mapreduce.Job: map 0% reduce 0%
18/10/06 22:04:10 INFO mapreduce.Job: map 100% reduce 0%
18/10/06 22:04:15 INFO mapreduce.Job: map 100% reduce 100%
18/10/06 22:04:15 INFO mapreduce.Job: Job job_1538279611801_0003 completed successfully
18/10/06 22:04:15 INFO mapreduce.Job: Counters: 49
    File System Counters
        FILE: Number of bytes read=218
        FILE: Number of bytes written=215817
        FILE: Number of read operations=0
        FILE: Number of large read operations=0
        FILE: Number of write operations=0
        HDFS: Number of bytes read=5852522
        HDFS: Number of bytes written=44
        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)=3380
        Total time spent by all reduces in occupied slots (ms)=2217
        Total time spent by all map tasks (ms)=3380
        Total time spent by all reduce tasks (ms)=2217
        Total vcore-milliseconds taken by all map tasks=3380
        Total vcore-milliseconds taken by all reduce tasks=2217
        Total megabyte-milliseconds taken by all map tasks=3461120
```

```
Applications Places System Sat Oct 6, 10:04 PM Acadgild
acadgild@localhost:~/Desktop
File Edit View Search Terminal Help
Reduce input groups=3
Reduce shuffle bytes=106
Reduce input records=6
Reduce output records=3
Spilled Records=18
Shuffled Maps =1
Failed Shuffles=0
Merged Map outputs=1
GC time elapsed (ms)=92
CPU time spent (ms)=3120
Physical memory (bytes) snapshot=425807872
Virtual memory (bytes) snapshot=4167393280
Total committed heap usage (bytes)=295174144
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=5852404
File Output Format Counters
Bytes Written=44
The standard deviation is: 2.2845207067745963
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost Desktop]$ hadoop fs -ls /outputassignment3.2
18/10/06 22:04:36 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 2018-10-06 22:04 /outputassignment3.2/_SUCCESS
-rw-r--r-- 1 acadgild supergroup 44 2018-10-06 22:04 /outputassignment3.2/part-r-00000
[acadgild@localhost Desktop]$ hadoop fs -cat /outputassignment3.2/part-r-00000
18/10/06 22:04:45 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
count 959893
length 4306887
square 24334031
[acadgild@localhost Desktop]$
```