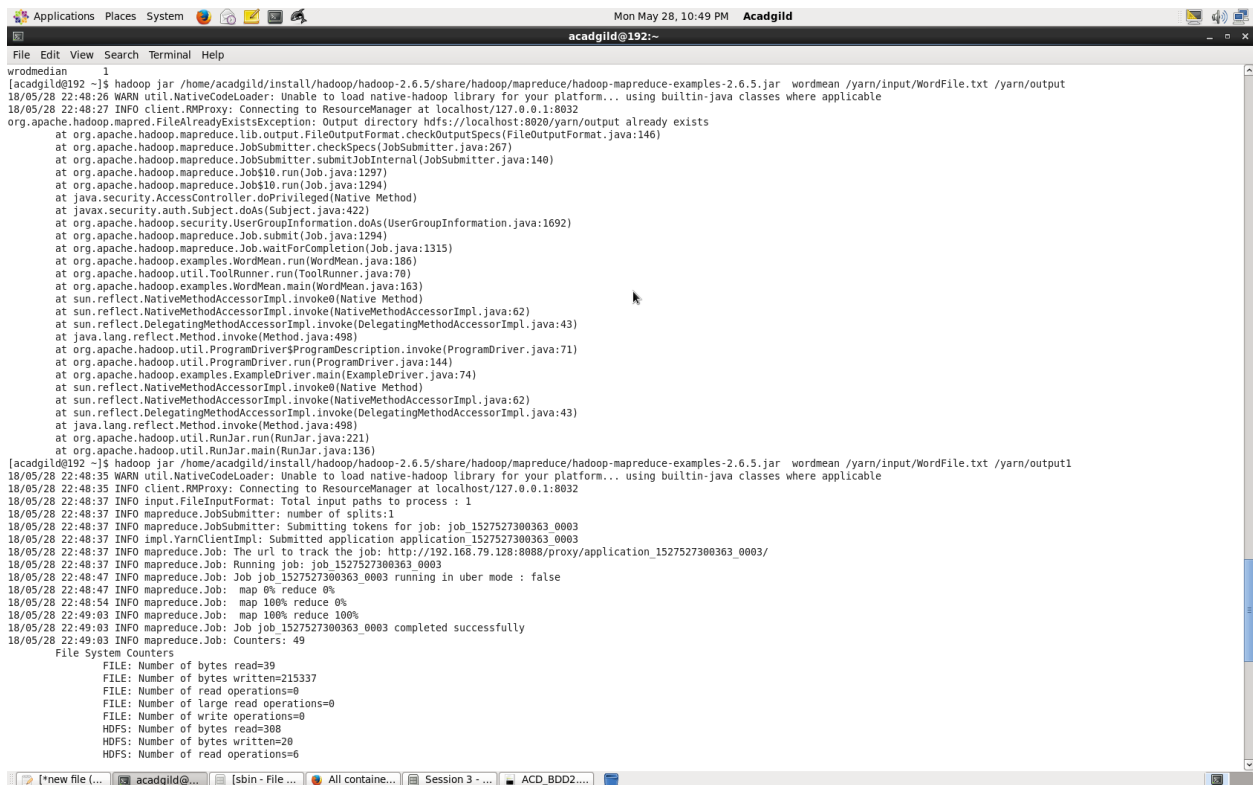


Task:1

Calculate word mean using hadoop-mapreduce-examples-2.6.5.jar

Here I have created one WordFile.txt using vi WordFile.txt from terminal & I have executed job for wordmean from terminal as below



```
wrodrmedian 1
[acagdild@0192 ~]$ hadoop jar /home/acagdild/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce/hadoop-mapreduce-examples-2.6.5.jar wordmean /yarn/input/WordFile.txt /yarn/output
18/05/28 22:48:26 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/05/28 22:48:27 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
org.apache.hadoop.mapred.FileAlreadyExistsException: Output directory hdfs://localhost:8020/yarn/output already exists
    at org.apache.hadoop.mapreduce.lib.output.FileOutputFormat.checkOutputSpecs(FileOutputFormat.java:146)
    at org.apache.hadoop.mapreduce.JobSubmitter.checkSpecs(JobSubmitter.java:267)
    at org.apache.hadoop.mapreduce.JobSubmitter.submitJobInternal(JobSubmitter.java:140)
    at org.apache.hadoop.mapreduce.JobS10.run(Job.java:1297)
    at org.apache.hadoop.mapreduce.JobS10.run(Job.java:1294)
    at java.security.AccessController.doPrivileged(Native Method)
    at javax.security.auth.Subject.doAs(Subject.java:422)
    at org.apache.hadoop.mapreduce.security.UserGroupInformation.doAs(UserGroupInformation.java:1692)
    at org.apache.hadoop.mapreduce.Job.submit(Job.java:1294)
    at org.apache.hadoop.mapreduce.Job.waitForCompletion(Job.java:1315)
    at org.apache.hadoop.examples.WordMean.run(WordMean.java:186)
    at org.apache.hadoop.util.ToolRunner.run(ToolRunner.java:70)
    at org.apache.hadoop.examples.WordMean.main(WordMean.java:163)
    at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
    at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:62)
    at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)
    at java.lang.reflect.Method.invoke(Method.java:498)
    at org.apache.hadoop.util.ProgramDriver$ProgramDescription.invoke(ProgramDriver.java:71)
    at org.apache.hadoop.util.ProgramDriver.run(ProgramDriver.java:144)
    at org.apache.hadoop.examples.ExampleDriver.main(ExampleDriver.java:74)
    at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
    at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:62)
    at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)
    at java.lang.reflect.Method.invoke(Method.java:498)
    at org.apache.hadoop.util.RunJar.run(RunJar.java:221)
    at org.apache.hadoop.util.RunJar.main(RunJar.java:136)
[acagdild@0192 ~]$ hadoop jar /home/acagdild/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce/hadoop-mapreduce-examples-2.6.5.jar wordmean /yarn/input/WordFile.txt /yarn/output1
18/05/28 22:48:35 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/05/28 22:48:35 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/05/28 22:48:37 INFO input.FileInputFormat: Total input paths to process : 1
18/05/28 22:48:37 INFO mapreduce.JobSubmitter: number of splits:1
18/05/28 22:48:37 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1527527300363_0003
18/05/28 22:48:37 INFO impl.YarnClientImpl: Submitted application application_1527527300363_0003
18/05/28 22:48:37 INFO mapreduce.Job: The url to track the job: http://192.168.79.128:8088/proxy/application_1527527300363_0003/
18/05/28 22:48:37 INFO mapreduce.Job: Running job: job_1527527300363_0003
18/05/28 22:48:47 INFO mapreduce.Job: Job job_1527527300363_0003 running in uber mode : false
18/05/28 22:48:47 INFO mapreduce.Job: map 0% reduce 0%
18/05/28 22:48:54 INFO mapreduce.Job: map 100% reduce 0%
18/05/28 22:49:03 INFO mapreduce.Job: map 100% reduce 100%
18/05/28 22:49:03 INFO mapreduce.Job: Job job_1527527300363_0003 completed successfully
18/05/28 22:49:03 INFO mapreduce.Job: Counters: 49
File System Counters
  FILE: Number of bytes read=39
  FILE: Number of bytes written=215337
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=308
  HDFS: Number of bytes written=20
  HDFS: Number of read operations=6
```

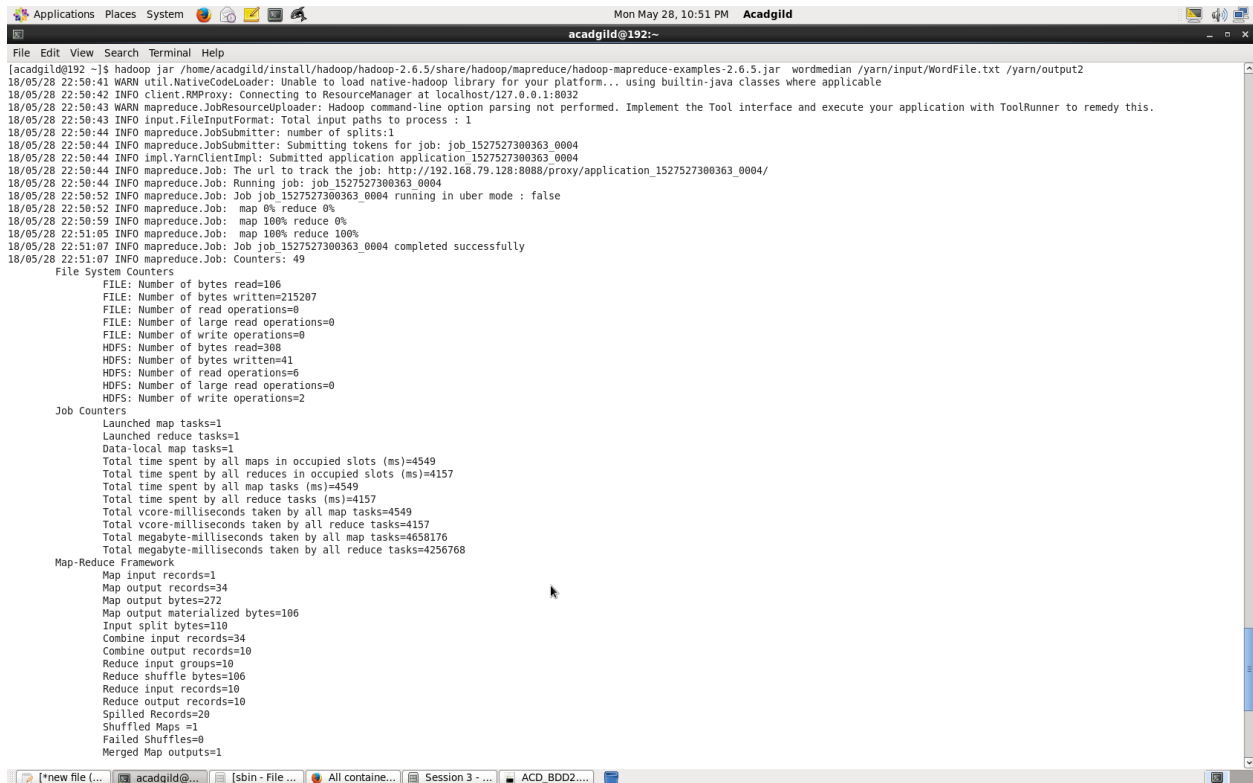
```
Applications Places System Mon May 28, 10:49 PM Acadgild
acadgild@192:~
File Edit View Search Terminal Help
18/05/28 22:49:03 INFO mapreduce.Job: Job job_1527527300363_0003 completed successfully
18/05/28 22:49:03 INFO mapreduce.Job: Counters: 49
File System Counters
  FILE: Number of bytes read=39
  FILE: Number of bytes written=215337
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=308
  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)=5610
  Total time spent by all reduces in occupied slots (ms)=5663
  Total time spent by all map tasks (ms)=5610
  Total time spent by all reduce tasks (ms)=5663
  Total vcore-milliseconds taken by all map tasks=5610
  Total vcore-milliseconds taken by all reduce tasks=5663
  Total megabyte-milliseconds taken by all map tasks=5744640
  Total megabyte-milliseconds taken by all reduce tasks=5798912
Map-Reduce Framework
  Map input records=1
  Map output records=68
  Map output bytes=986
  Map output materialized bytes=39
  Input split bytes=110
  Combine input records=68
  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)=193
  CPU time spent (ms)=1620
  Physical memory (bytes) snapshot=328122368
  Virtual memory (bytes) snapshot=4126875648
  Total committed heap usage (bytes)=238424576
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=198
```

Please find the answer for task 1

```
Applications Places System Mon May 28, 10:49 PM Acadgild
acadgild@192:~
File Edit View Search Terminal Help
  HDFS: Number of bytes read=308
  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)=5610
  Total time spent by all reduces in occupied slots (ms)=5663
  Total time spent by all map tasks (ms)=5610
  Total time spent by all reduce tasks (ms)=5663
  Total vcore-milliseconds taken by all map tasks=5610
  Total vcore-milliseconds taken by all reduce tasks=5663
  Total megabyte-milliseconds taken by all map tasks=5744640
  Total megabyte-milliseconds taken by all reduce tasks=5798912
Map-Reduce Framework
  Map input records=1
  Map output records=68
  Map output bytes=986
  Map output materialized bytes=39
  Input split bytes=110
  Combine input records=68
  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)=193
  CPU time spent (ms)=1620
  Physical memory (bytes) snapshot=328122368
  Virtual memory (bytes) snapshot=4126875648
  Total committed heap usage (bytes)=238424576
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=198
File Output Format Counters
  Bytes Written=20
The mean is: 4.764705882352941
[acadgild@192 ~]$ hadoop fs -cat /yarn/output1/part-r-00000
18/05/28 22:49:28 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
count      34
length    162
[acadgild@192 ~]$
```

Task:2

Calculate word median using hadoop-mapreduce-examples-2.6.5.jar



```
acacgild@192:~$ hadoop jar /home/acacgild/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce/hadoop-mapreduce-examples-2.6.5.jar wordmedian /yarn/input/WordFile.txt /yarn/output2
18/05/28 22:50:41 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/05/28 22:50:42 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/05/28 22:50:43 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
18/05/28 22:50:43 INFO input.FileInputFormat: Total input paths to process : 1
18/05/28 22:50:44 INFO mapreduce.JobSubmitter: number of splits:1
18/05/28 22:50:44 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1527527300363_0004
18/05/28 22:50:44 INFO impl.YarnClientImpl: Submitted application application_1527527300363_0004
18/05/28 22:50:44 INFO mapreduce.Job: The url to track the job: http://192.168.79.128:8088/proxy/application_1527527300363_0004/
18/05/28 22:50:44 INFO mapreduce.Job: Running job: job_1527527300363_0004
18/05/28 22:50:52 INFO mapreduce.Job: Job job_1527527300363_0004 running in uber mode : false
18/05/28 22:50:52 INFO mapreduce.Job: map 0% reduce 0%
18/05/28 22:50:59 INFO mapreduce.Job: map 100% reduce 0%
18/05/28 22:51:05 INFO mapreduce.Job: map 100% reduce 100%
18/05/28 22:51:07 INFO mapreduce.Job: Job job_1527527300363_0004 completed successfully
18/05/28 22:51:07 INFO mapreduce.Job: Counters: 49
File System Counters
  FILE: Number of bytes read=106
  FILE: Number of bytes written=215207
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=308
  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)=4549
  Total time spent by all reduces in occupied slots (ms)=4157
  Total time spent by all map tasks (ms)=4549
  Total time spent by all reduce tasks (ms)=4157
  Total vcore-millisecods taken by all map tasks=4549
  Total vcore-millisecods taken by all reduce tasks=4157
  Total megabyte-millisecods taken by all map tasks=4658176
  Total megabyte-millisecods taken by all reduce tasks=4256768
Map-Reduce Framework
  Map input records=1
  Map output records=34
  Map output bytes=272
  Map output materialized bytes=106
  Input split bytes=110
  Combine input records=34
  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
```

Please find the answer for Task2

```
Applications Places System Mon May 28, 10:51 PM Acadgild
acadmild@192:~
File Edit View Search Terminal Help
Data-local map tasks=1
Total time spent by all maps in occupied slots (ms)=4549
Total time spent by all reduces in occupied slots (ms)=4157
Total time spent by all map tasks (ms)=4549
Total time spent by all reduce tasks (ms)=4157
Total vcore-milliseconds taken by all map tasks=4549
Total vcore-milliseconds taken by all reduce tasks=4157
Total megabyte-milliseconds taken by all map tasks=4658176
Total megabyte-milliseconds taken by all reduce tasks=4256768
Map-Reduce Framework
Map input records=1
Map output records=34
Map output bytes=272
Map output materialized bytes=106
Input split bytes=110
Combine input records=34
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)=120
CPU time spent (ms)=1220
Physical memory (bytes) snapshot=326651904
Virtual memory (bytes) snapshot=4126867456
Total committed heap usage (bytes)=238424576
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=198
File Output Format Counters
Bytes Written=41
The median is: 4
[acadmild@192 ~]$ hadoop fs -cat /yarn/output2/part-r-000000
18/05/28 22:51:20 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
1 3
2 5
3 3
4 9
5 2
6 2
7 4
8 2
9 2
10 2
[acadmild@192 ~]$
```

Task:3

Calculate word standard deviation using hadoop-mapreduce-examples-2.6.5.jar

```
Applications Places System Mon May 28, 10:53 PM acadgild
acadgild@192:~
File Edit View Search Terminal Help
[acadgild@192 ~]$ hadoop jar /home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce/hadoop-mapreduce-examples-2.6.5.jar wordstandarddeviation /yarn/input/WordFile.txt /yarn/output3
18/05/28 22:52:38 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/05/28 22:52:39 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/05/28 22:52:40 INFO input.FileInputFormat: Total input paths to process : 1
18/05/28 22:52:40 INFO mapreduce.JobSubmitter: number of splits:1
18/05/28 22:52:40 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1527527300363_0005
18/05/28 22:52:41 INFO impl.YarnClientImpl: Submitted application application_1527527300363_0005
18/05/28 22:52:41 INFO mapreduce.Job: The url to track the job: http://192.168.79.128:8088/proxy/application_1527527300363_0005/
18/05/28 22:52:41 INFO mapreduce.Job: Running job: job_1527527300363_0005
18/05/28 22:52:49 INFO mapreduce.Job: Job job_1527527300363_0005 running in uber mode : false
18/05/28 22:52:49 INFO mapreduce.Job: map 0% reduce 0%
18/05/28 22:52:54 INFO mapreduce.Job: map 100% reduce 0%
18/05/28 22:53:00 INFO mapreduce.Job: map 100% reduce 100%
18/05/28 22:53:01 INFO mapreduce.Job: Job job_1527527300363_0005 completed successfully
18/05/28 22:53:02 INFO mapreduce.Job: Counters: 49
File System Counters
  FILE: Number of bytes read=56
  FILE: Number of bytes written=215531
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=308
  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
  Launched reduce tasks=1
  Data-local map tasks=1
  Total time spent by all maps in occupied slots (ms)=3485
  Total time spent by all reduces in occupied slots (ms)=3886
  Total time spent by all map tasks (ms)=3485
  Total time spent by all reduce tasks (ms)=3886
  Total vcore-milliseconds taken by all map tasks=3485
  Total vcore-milliseconds taken by all reduce tasks=3886
  Total megabyte-milliseconds taken by all map tasks=3568640
  Total megabyte-milliseconds taken by all reduce tasks=3979264
Map-Reduce Framework
  Map input records=1
  Map output records=102
  Map output bytes=1496
  Map output materialized bytes=56
  Input split bytes=110
  Combine input records=102
  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)=119
  CPU time spent (ms)=1190
  Physical memory (bytes) snapshot=326836224
  Virtual memory (bytes) snapshot=4126875648
  Total committed heap usage (bytes)=238424576
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=198
File Output Format Counters
  Bytes Written=32
The standard deviation is: 2.6015699279407834
[acadgild@192 ~]$ hadoop fs -cat /yarn/output3/part-r-00000
count      34
length     162
square     1002
[acadgild@192 ~]$
```

Please find the answer for Task 3

```
Applications Places System Mon May 28, 10:53 PM acadgild
acadgild@192:~
File Edit View Search Terminal Help
HDFS: Number of bytes written=32
HDFS: Number of read operations=0
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)=3485
  Total time spent by all reduces in occupied slots (ms)=3886
  Total time spent by all map tasks (ms)=3485
  Total time spent by all reduce tasks (ms)=3886
  Total vcore-milliseconds taken by all map tasks=3485
  Total vcore-milliseconds taken by all reduce tasks=3886
  Total megabyte-milliseconds taken by all map tasks=3568640
  Total megabyte-milliseconds taken by all reduce tasks=3979264
Map-Reduce Framework
  Map input records=1
  Map output records=102
  Map output bytes=1496
  Map output materialized bytes=56
  Input split bytes=110
  Combine input records=102
  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)=119
  CPU time spent (ms)=1190
  Physical memory (bytes) snapshot=326836224
  Virtual memory (bytes) snapshot=4126875648
  Total committed heap usage (bytes)=238424576
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=198
File Output Format Counters
  Bytes Written=32
The standard deviation is: 2.6015699279407834
[acadgild@192 ~]$ hadoop fs -cat /yarn/output3/part-r-00000
18/05/28 22:53:16 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
count      34
length     162
square     1002
[acadgild@192 ~]$
```