

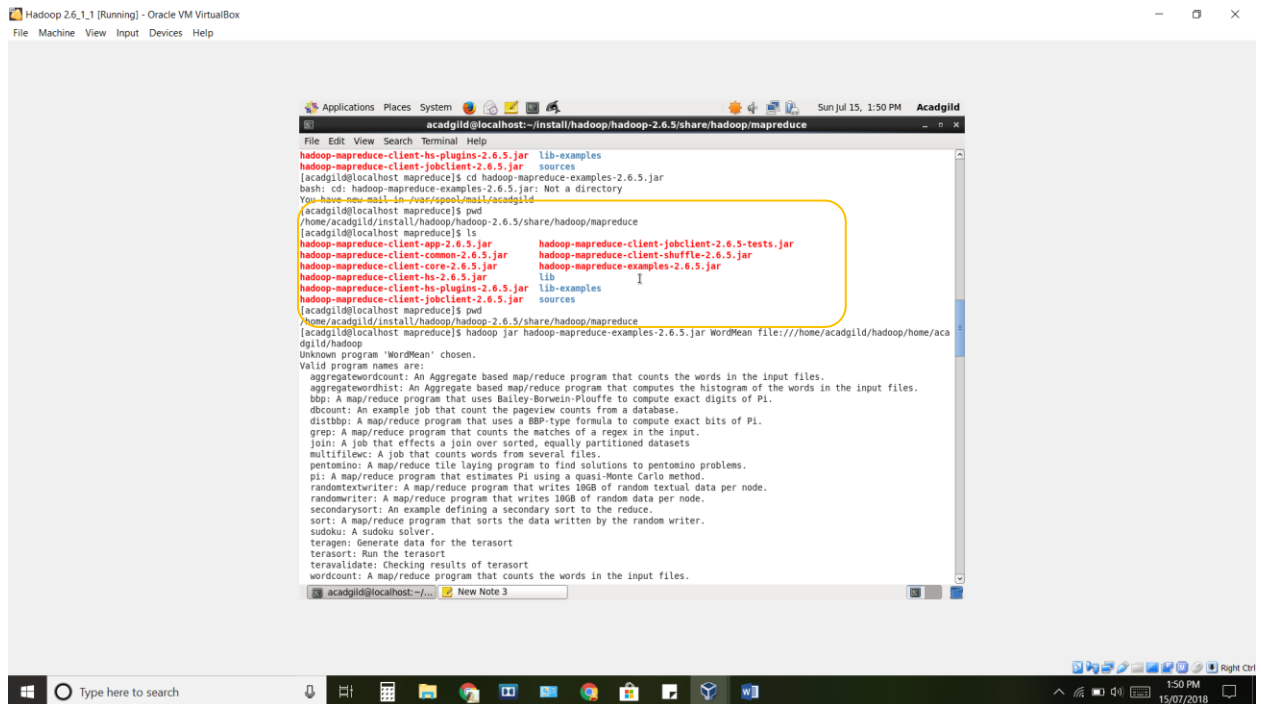
Assignment 3

Tasks of Big DATA Hadoop Development

- Execute Programs using Jar Hadoop-mapreduce-examples-2.6.5.jar

1. WordMean

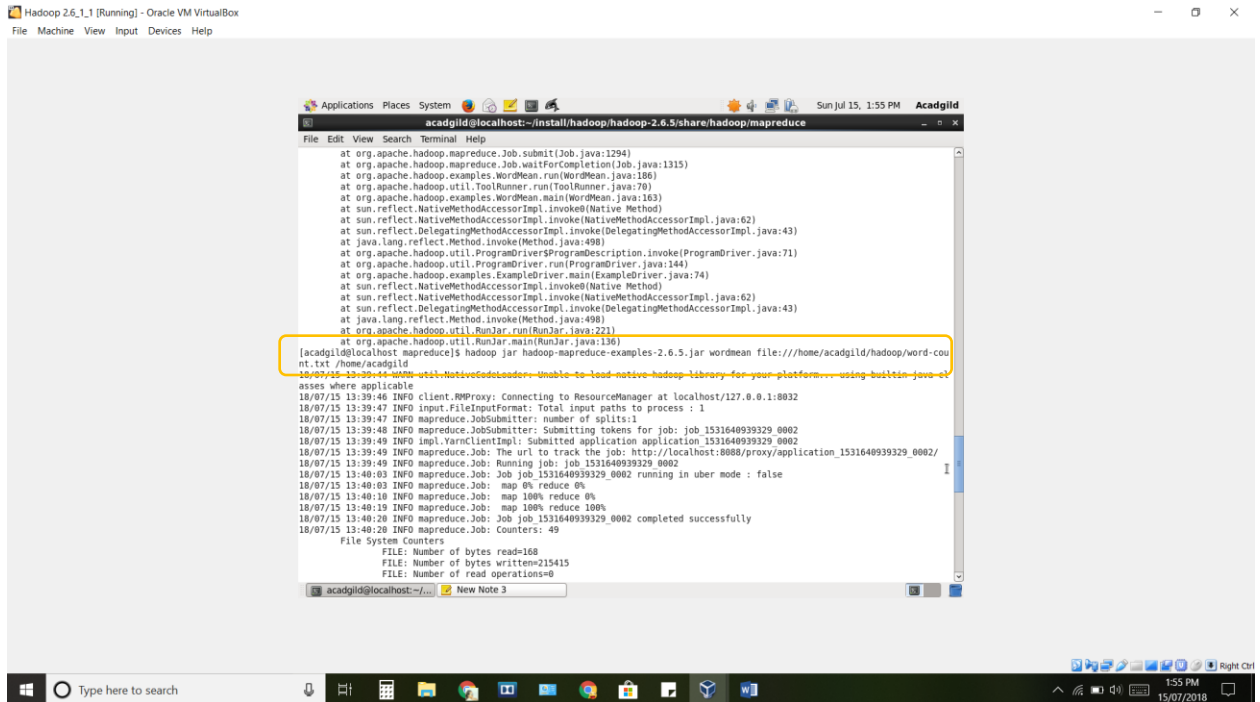
- Go to path where jar file is found



The screenshot shows a terminal window titled "Hadoop 2.6.1_1 [Running] - Oracle VM VirtualBox". The terminal is running a series of commands to set up the environment and execute the WordMean program. The commands and their outputs are as follows:

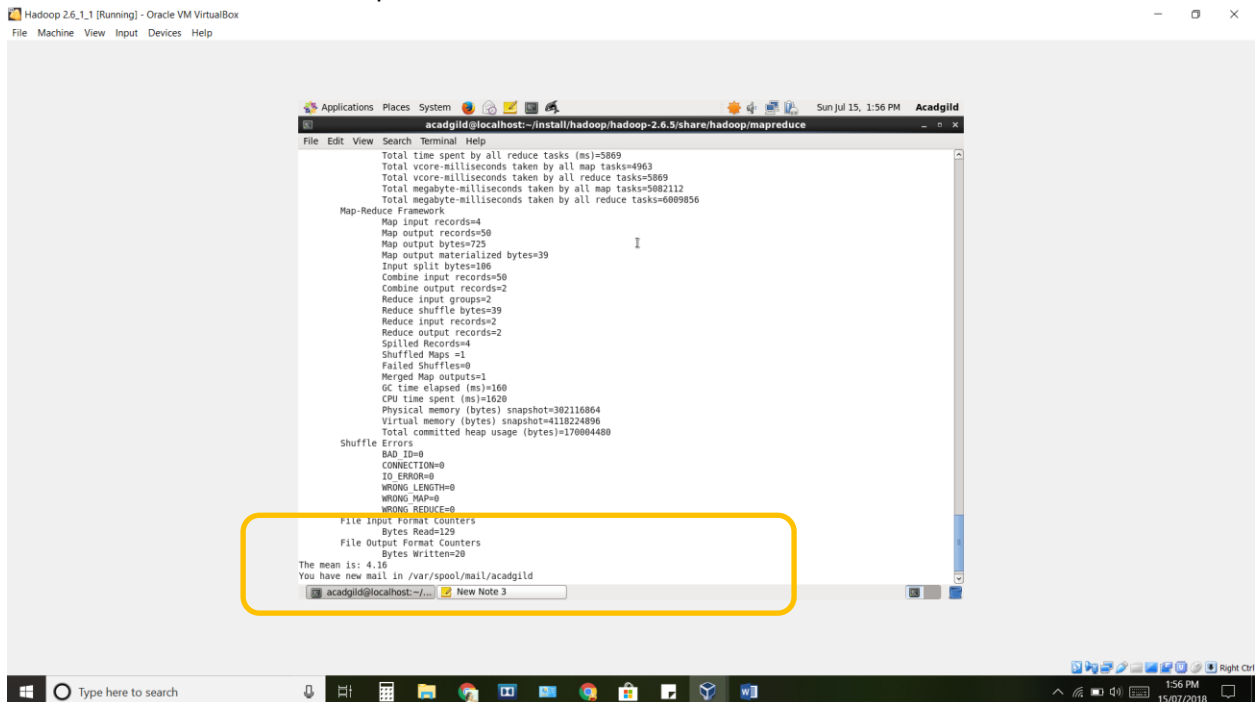
```
acacgild@localhost:~/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce
File Edit View Search Terminal Help
hadoop-mapreduce-client-hs-plugins-2.6.5.jar lib-examples
hadoop-mapreduce-client-jobclient-2.6.5.jar sources
[acacgild@localhost mapreduce]$ cd hadoop-mapreduce-examples-2.6.5.jar
bash: cd: hadoop-mapreduce-examples-2.6.5.jar: Not a directory
You have new mail in /var/spool/mail/acacgild
[acacgild@localhost mapreduce]$ pwd
/home/acacgild/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce
[acacgild@localhost mapreduce]$ ls
hadoop-mapreduce-client-common-2.6.5.jar  hadoop-mapreduce-client-jobclient-2.6.5-tests.jar
hadoop-mapreduce-client-core-2.6.5.jar    hadoop-mapreduce-client-shuffle-2.6.5.jar
hadoop-mapreduce-client-hs-2.6.5.jar      hadoop-mapreduce-examples-2.6.5.jar
hadoop-mapreduce-client-hs-plugins-2.6.5.jar lib-examples
hadoop-mapreduce-client-jobclient-2.6.5.jar sources
[acacgild@localhost mapreduce]$ pwd
/home/acacgild/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce
[acacgild@localhost mapreduce]$ hadoop jar hadoop-mapreduce-examples-2.6.5.jar WordMean file:///home/acacgild/hadoop/home/acacgild/hadoop
Unknown program 'WordMean' chosen.
Valid program names are:
aggregatewordcount: An Aggregate based map/reduce program that counts the words in the input files.
aggregatewordhist: An Aggregate based map/reduce program that computes the histogram of the words in the input files.
bbp: A map/reduce program that uses Bailey-Borwein-Plouffe to compute exact digits of Pi.
dbcount: An example job that count the pageview counts from a database.
distbbp: A map/reduce program that uses a BBP-type formula to compute exact bits of Pi.
grep: A map/reduce program that counts the matches of a regex in the input.
join: A job that effects a join over sorted, equally partitioned datasets
multifiles: A job that counts words from several files.
pentomino: A map/reduce tile laying program to find solutions to pentomino problems.
pi: A map/reduce program that estimates Pi using a quasi-Monte Carlo method.
randomtextwriter: A map/reduce program that writes 10GB of random textual data per node.
randomwriter: A map/reduce program that writes 10GB of random data per node.
secondarysort: An example defining a secondary sort to the reduce.
sort: A map/reduce program that sorts the data written by the random writer.
sudoku: A sudoku solver.
terasgen: Generate data for the terasort
terasort: Run the terasort
teravalidate: Checking results of terasort
wordcount: A map/reduce program that counts the words in the input files.
```

ii. Execute command to Calculate Word Mean on file word-count.txt



```
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.invoke(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$ProgramDescription.invoke(ProgramDriver.java:71)
at org.apache.hadoop.util.ProgramDriver.run(ProgramDriver.java:144)
at org.apache.hadoop.examples.ExampleDriver.main(ExampleDriver.java:74)
at sun.reflect.NativeMethodAccessorImpl.invoke(Native Method)
at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:62)
at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)
at java.lang.reflect.Method.invoke(Method.java:498)
at org.apache.hadoop.util.RunJar.run(RunJar.java:221)
at org.apache.hadoop.util.RunJar.main(RunJar.java:136)
[acadgild@localhost ~]$ hadoop jar hadoop-mapreduce-examples-2.6.5.jar wordmean file:///home/acadgild/hadoop/word-count.txt /home/acadgild
18/07/15 13:39:44 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java-based
asses where applicable
18/07/15 13:39:46 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/07/15 13:39:47 INFO input.FileInputFormat: Total input paths to process : 1
18/07/15 13:39:47 INFO mapreduce.JobSubmitter: number of splits:1
18/07/15 13:39:48 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1531640939329_0002
18/07/15 13:39:49 INFO impl.YarnClientImpl: Submitted application application_1531640939329_0002
18/07/15 13:39:49 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1531640939329_0002/
18/07/15 13:39:49 INFO mapreduce.Job: Running job: job_1531640939329_0002
18/07/15 13:40:03 INFO mapreduce.Job: Job job_1531640939329_0002 running in uber mode : false
18/07/15 13:40:03 INFO mapreduce.Job: map 0% reduce 0%
18/07/15 13:40:10 INFO mapreduce.Job: map 100% reduce 0%
18/07/15 13:40:19 INFO mapreduce.Job: map 100% reduce 100%
18/07/15 13:40:20 INFO mapreduce.Job: Job job_1531640939329_0002 completed successfully
18/07/15 13:40:20 INFO mapreduce.Job: Counters: 49
File System Counters
  FILE: Number of bytes read=168
  FILE: Number of bytes written=215415
  FILE: Number of read operations=0
```

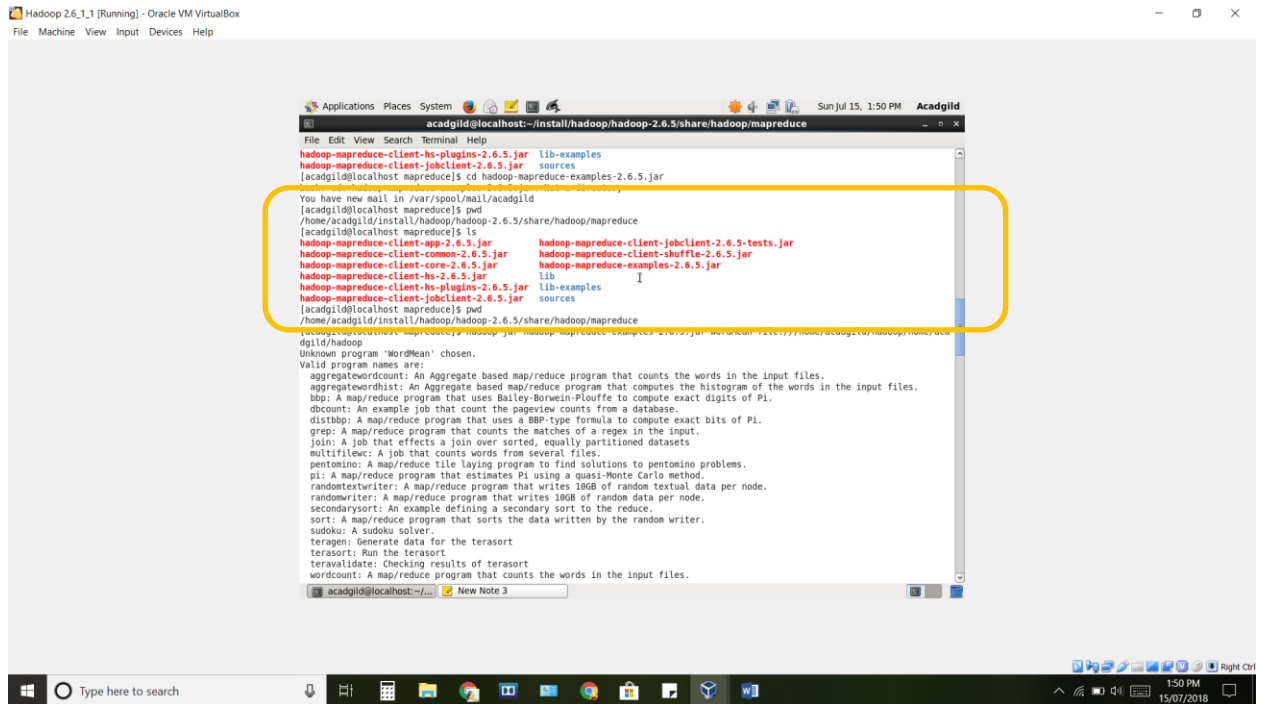
iii. Final WordMean Output



```
Total time spent by all reduce tasks (ms)=5869
Total vcore-milliseconds taken by all map tasks=4963
Total vcore-milliseconds taken by all reduce tasks=5869
Total megabyte-milliseconds taken by all map tasks=5082112
Total megabyte-milliseconds taken by all reduce tasks=6009856
Map-Reduce Framework
  Map input records=4
  Map output records=50
  Map output bytes=725
  Map output materialized bytes=39
  Input split bytes=106
  Combine input records=50
  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)=160
  CPU time spent (ms)=1620
  Physical memory (bytes) snapshot=302116864
  Virtual memory (bytes) snapshot=4118224896
  Total committed heap usage (bytes)=170004488
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=129
File Output Format Counters
  Bytes Written=20
The mean is: 4.16
You have new mail in /var/spool/mail/acadgild
```

2. WordMedian

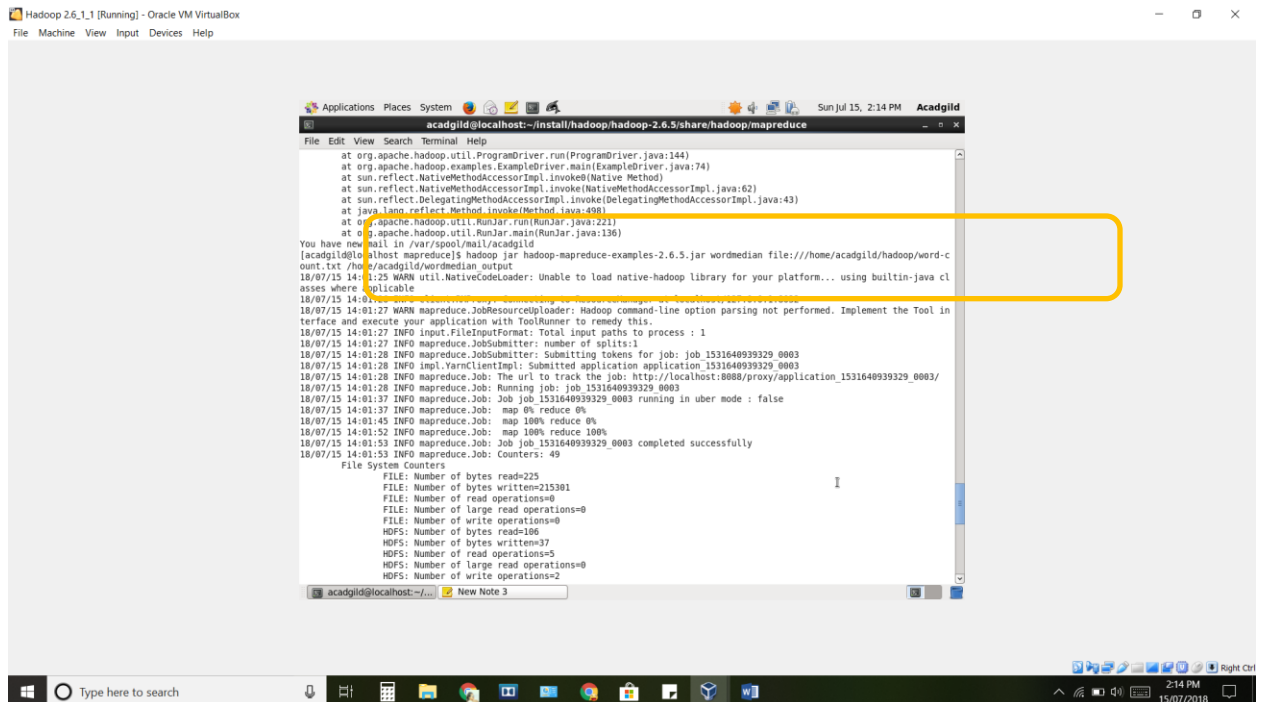
i. Go to path where Jar file is found



```
Hadoop 2.6.1_1 [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help

acacgild@localhost: ~/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce
File Edit View Search Terminal Help
hadoop-mapreduce-client-hs-plugins-2.6.5.jar lib-examples
hadoop-mapreduce-client-jobclient-2.6.5.jar sources
[acacgild@localhost mapreduce]$ cd /home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce
You have new mail in /var/spool/mail/acacgild
[acacgild@localhost mapreduce]$ pwd
/home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce
[acacgild@localhost mapreduce]$ ls
hadoop-mapreduce-client-app-2.6.5.jar      hadoop-mapreduce-client-jobclient-2.6.5-tests.jar
hadoop-mapreduce-client-common-2.6.5.jar   hadoop-mapreduce-client-shuffle-2.6.5.jar
hadoop-mapreduce-client-core-2.6.5.jar     hadoop-mapreduce-examples-2.6.5.jar
hadoop-mapreduce-client-hs-2.6.5.jar       lib
hadoop-mapreduce-client-hs-plugins-2.6.5.jar lib-examples
hadoop-mapreduce-client-jobclient-2.6.5.jar sources
[acacgild@localhost mapreduce]$ pwd
/home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce
[acacgild@localhost mapreduce]$ hadoop jar hadoop-mapreduce-examples-2.6.5.jar wordcount word-count.txt
Unknown program 'wordcount' chosen.
Valid program names are:
aggregatewordcount: An Aggregate based map/reduce program that counts the words in the input files.
aggregatetextcount: An Aggregate based map/reduce program that counts the histogram of the words in the input files.
bbp: A map/reduce program that uses Bailey-Borwein-Plouffe to compute exact digits of Pi.
dbcount: An example job that counts the pageview counts from a database.
distbbp: A map/reduce program that uses a BBP-type formula to compute exact bits of Pi.
grep: A map/reduce program that counts the matches of a regex in the input.
join: A job that effects a join over sorted, equally partitioned datasets
multifiles: A job that counts words from several files.
pentomino: A map/reduce tile laying program to find solutions to pentomino problems.
pi: A map/reduce program that estimates Pi using a quasi-Monte Carlo method.
randomtextwriter: A map/reduce program that writes 1000 of random textual data per node.
randomwriter: A map/reduce program that writes 1000 of random data per node.
secondarysort: An example defining a secondary sort to the reduce.
sort: A map/reduce program that sorts the data written by the random writer.
sudoku: A sudoku solver.
terasort: Generate data for the terasort
terasort: Run the terasort
terasort: Checking results of terasort
wordcount: A map/reduce program that counts the words in the input files.
```

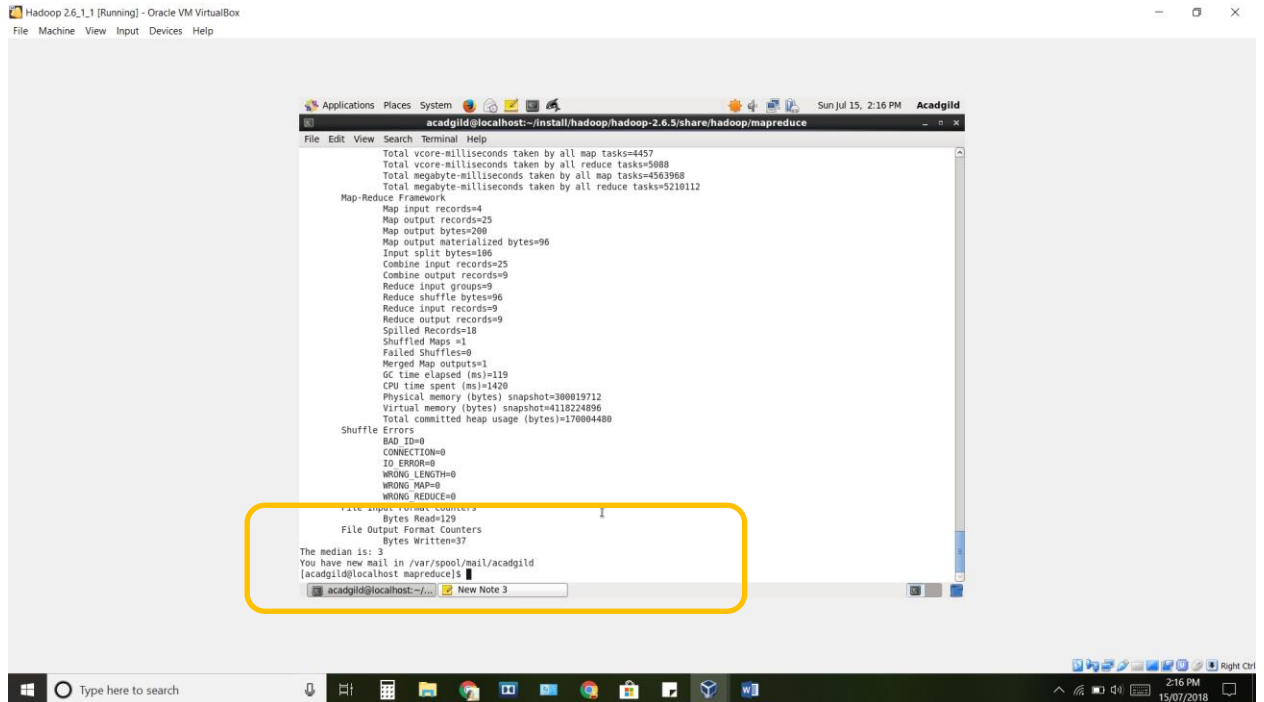
ii. Execute command to Calculate Word Median on file word-count.txt



```
Hadoop 2.6.1_1 [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help

acacgild@localhost: ~/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce
File Edit View Search Terminal Help
at org.apache.hadoop.util.ProgramDriver.run(ProgramDriver.java:144)
at org.apache.hadoop.examples.ExampleDriver.main(ExampleDriver.java:74)
at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethod)
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)
You have new mail in /var/spool/mail/acacgild
[acacgild@localhost mapreduce]$ hadoop jar hadoop-mapreduce-examples-2.6.5.jar wordmedian file:///home/acadgild/hadoop/word-count.txt /home/acadgild/wordmedian output
18/07/15 14:01:25 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/07/15 14:01:27 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool in interface and execute your application with ToolRunner to remedy this.
18/07/15 14:01:27 INFO InputFileInputFormat: Total input paths to process : 1
18/07/15 14:01:27 INFO mapreduce.JobSubmitter: number of splits:1
18/07/15 14:01:28 INFO impl.VanClientImpl: Submitted application for job: job_1531640939329_0003
18/07/15 14:01:28 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1531640939329_0003/
18/07/15 14:01:28 INFO mapreduce.Job: Running job: job_1531640939329_0003
18/07/15 14:01:37 INFO mapreduce.Job: Job job_1531640939329_0003 running in uber mode : false
18/07/15 14:01:45 INFO mapreduce.Job: map 0% reduce 0%
18/07/15 14:01:52 INFO mapreduce.Job: map 100% reduce 100%
18/07/15 14:01:53 INFO mapreduce.Job: Job job_1531640939329_0003 completed successfully
18/07/15 14:01:53 INFO mapreduce.Job: Counters: 49
File System Counters
FILE: Number of bytes read=225
FILE: Number of bytes written=215301
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=106
HDFS: Number of bytes written=37
HDFS: Number of read operations=5
HDFS: Number of large read operations=0
HDFS: Number of write operations=2
```

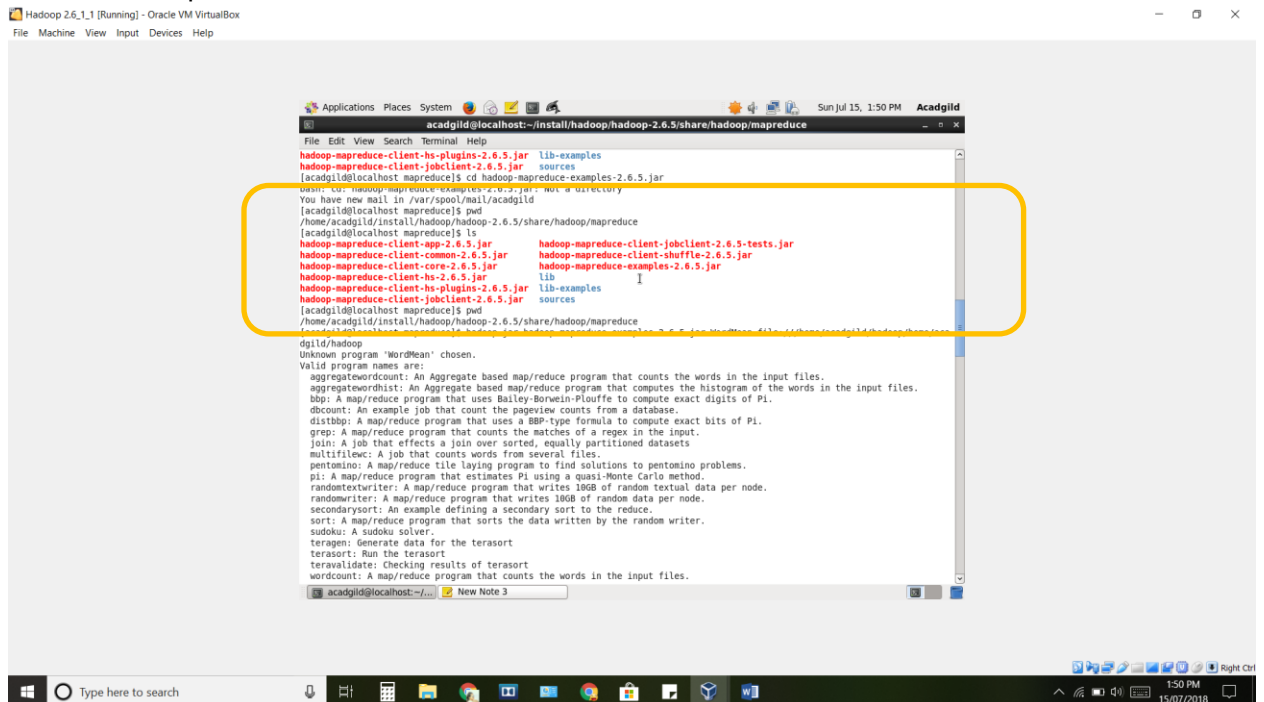
iii. Final Output of WordMedian



```
acacgild@localhost:~/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce$ cat /dev/null
Total vcore-millisecs taken by all map tasks=4457
Total vcore-millisecs taken by all reduce tasks=5088
Total megabyte-millisecs taken by all map tasks=4563968
Total megabyte-millisecs taken by all reduce tasks=5210112
Map-Reduce Framework
  Map input records=4
  Map output records=25
  Map output bytes=208
  Map output materialized bytes=96
  Input split bytes=186
  Combine input records=25
  Combine output records=9
  Reduce input groups=9
  Reduce shuffle bytes=96
  Reduce input records=9
  Reduce output records=9
  Spilled Records=18
  Shuffled Maps=1
  Failed Shuffles=0
  Merged Map outputs=1
  GC time elapsed (ms)=119
  CPU time spent (ms)=1420
  Physical memory (bytes) snapshot=306019712
  Virtual memory (bytes) snapshot=118224096
  Total committed heap usage (bytes)=17004488
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=129
File Output Format Counters
  Bytes Written=37
The median is: 3
You have new mail in /var/spool/mail/acacgild
(acacgild@localhost mapreduce)$
```

3. Word Standard Deviation

i. Go to path where Jar file is found



```
acacgild@localhost:~/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce$ ls
hadoop-mapreduce-client-app-2.6.5.jar      hadoop-mapreduce-client-jobclient-2.6.5-tests.jar
hadoop-mapreduce-client-common-2.6.5.jar    hadoop-mapreduce-client-shuffle-2.6.5.jar
hadoop-mapreduce-client-core-2.6.5.jar      hadoop-mapreduce-examples-2.6.5.jar
hadoop-mapreduce-client-hs-2.6.5.jar        lib
hadoop-mapreduce-client-hs-plugins-2.6.5.jar lib-examples
hadoop-mapreduce-client-jobclient-2.6.5.jar sources
acacgild@localhost:~/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce$
```

ii. Execute command to Calculate WordStandardDeviation on file word-count.txt

```
acacgild@localhost:~/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce
File Edit View Search Terminal Help

Failed Shuffles=0
Merged Map outputs=1
GC time elapsed (ms)=119
CPU time spent (ms)=1429
Physical memory (bytes) snapshot=300019712
Virtual memory (bytes) snapshot=418224896
Total committed heap usage (bytes)=170004488

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=129
Bytes Written=37
The median is: 3
You have new mail in /var/spool/mail/acacgild
[acacgild@localhost mapreduce]$ hadoop jar hadoop-mapreduce-examples-2.6.5.jar wordstandarddeviation file:///home/acacgild/hadoop/word-count.txt /home/acacgild/wordsd_output
18/07/15 14:17:47 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/07/15 14:17:48 INFO client.RMProxy: Connection to ResourceManager at localhost/127.0.0.1:8032
18/07/15 14:17:50 INFO input.FileInputFormat: Total input paths to process : 1
18/07/15 14:17:50 INFO mapreduce.JobSubmitter: number of splits:1
18/07/15 14:17:50 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1531640939329_0004
18/07/15 14:17:50 INFO mapreduce.Job: The url to track the job: http://localhost:8080/proxy/application_1531640939329_0004/
18/07/15 14:17:50 INFO mapreduce.Job: Running job: job_1531640939329_0004
18/07/15 14:18:00 INFO mapreduce.Job: Job job_1531640939329_0004 running in uber mode : false
18/07/15 14:18:00 INFO mapreduce.Job: map 0% reduce 0%
18/07/15 14:18:07 INFO mapreduce.Job: map 100% reduce 0%
18/07/15 14:18:17 INFO mapreduce.Job: map 100% reduce 100%
18/07/15 14:18:17 INFO mapreduce.Job: Job job_1531640939329_0004 completed successfully
18/07/15 14:18:17 INFO mapreduce.Job: Counters: 49
File System Counters
FILE: Number of bytes read=185
acacgild@localhost:~/... New Note 3
```

iii. Final Output of WordStandardDeviation

```
acacgild@localhost:~/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce
File Edit View Search Terminal Help

Total vcore-millisecods taken by all map tasks=4797
Total vcore-millisecods taken by all reduce tasks=6717
Total megabyte-millisecods taken by all map tasks=4912128
Total megabyte-millisecods taken by all reduce tasks=6878208

Map-Reduce Framework
Map input records=4
Map output records=75
Map output bytes=1180
Map output materialized bytes=56
Input split bytes=106
Combine input records=75
Combine output records=3
Reduce input groups=3
Reduce shuffle bytes=56
Reduce input records=3
Reduce output records=3
Spilled Records=6
Failed Maps=1
Failed Shuffles=0
Merged Map outputs=1
GC time elapsed (ms)=156
CPU time spent (ms)=1940
Physical memory (bytes) snapshot=298164224
Virtual memory (bytes) snapshot=418216704
Total committed heap usage (bytes)=170004488

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=129
File Output Format Counters
Bytes Written=31
The standard deviation is: 2.6333248945012455
You have new mail in /var/spool/mail/acacgild
[acacgild@localhost mapreduce]$
acacgild@localhost:~/... New Note 3
```

