

Assignment 3.1

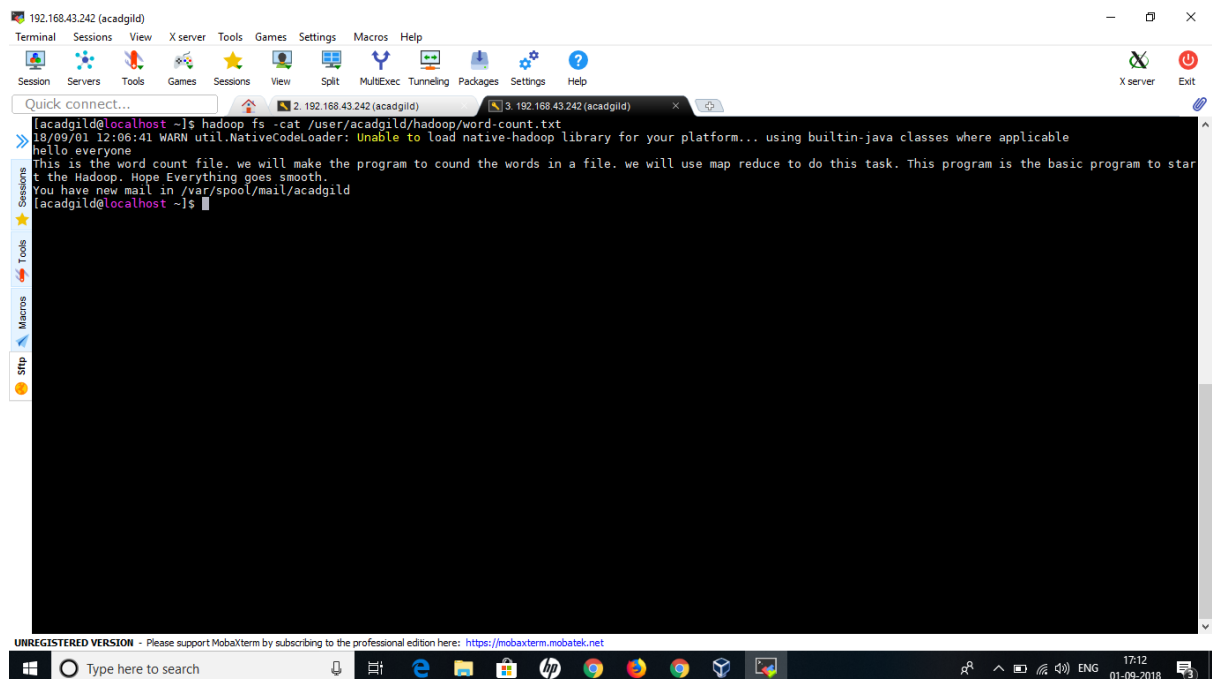
Task: Execute Word Median, Word Mean and Word Standard Deviation Programs using Jar file present in our VM.

1. **Word Median:** A map/reduce program that counts the median length of the words in the input files.

Command: `hadoop jar hadoop-mapreduce-examples-2.6.5.jar wordmedian/user/acadgild/hadoop/word-count.txt wordmedian`

Explanation: We have used the map reduce jar file and mentioned the name of the program as wordmedian then input file path as well as provide the output Path.

Input File Content :

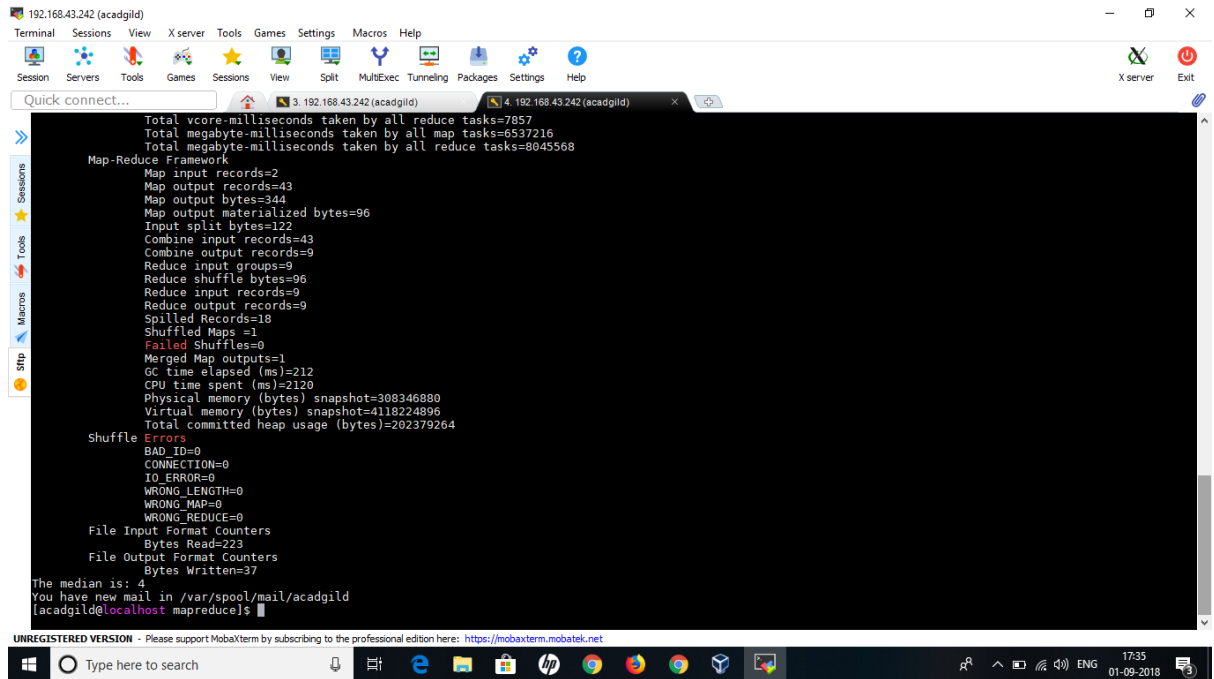


```
192.168.43.242 (acadgild)
Terminal Sessions View X server Tools Games Settings Macros Help
Session Servers Tools Games Sessions View Split MultiExec Tunneling Packages Settings Help
Quick connect...
2. 192.168.43.242 (acadgild) 3. 192.168.43.242 (acadgild)
[acadgild@localhost ~]$ hadoop fs -cat /user/acadgild/hadoop/word-count.txt
18/09/01 12:06:41 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
hello everyone
This is the word count file. we will make the program to count the words in a file. we will use map reduce to do this task. This program is the basic program to start the Hadoop. Hope Everything goes smooth.
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$
```

Command Execution:

```
192.168.43.242 (acadgild)
Terminal Sessions View X server Tools Games Settings Macros Help
Quick connect... 3 192.168.43.242 (acadgild) 4 192.168.43.242 (acadgild)
[acadgild@localhost ~]$ cd share/
[acadgild@localhost share]$ ls
doc hadoop
[acadgild@localhost share]$ cd hadoop/
[acadgild@localhost hadoop]$ ls
common hdfs httpfs kms mapreduce tools yarn
[acadgild@localhost hadoop]$ cd mapreduce/
[acadgild@localhost mapreduce]$ ls
hadoop-mapreduce-client-app-2.6.5.jar      hadoop-mapreduce-client-hs-2.6.5.jar      hadoop-mapreduce-client-jobclient-2.6.5-tests.jar  lib
hadoop-mapreduce-client-common-2.6.5.jar   hadoop-mapreduce-client-hs-plugins-2.6.5.jar  hadoop-mapreduce-client-shuffle-2.6.5.jar          lib-examples
hadoop-mapreduce-client-core-2.6.5.jar     hadoop-mapreduce-client-jobclient-2.6.5.jar  hadoop-mapreduce-examples-2.6.5.jar              sources
[acadgild@localhost mapreduce]$
[acadgild@localhost mapreduce]$ hadoop jar hadoop-mapreduce-examples-2.6.5.jar wordmedian /user/acadgild/hadoop/word-count.txt wordmedian
18/09/01 12:16:09 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/09/01 12:16:11 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/09/01 12:16:12 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with
  -D ToolRunner to remedy this.
18/09/01 12:16:13 INFO input.FileInputFormat: Total input paths to process : 1
18/09/01 12:16:13 INFO mapreduce.JobSubmitter: number of splits:1
18/09/01 12:16:13 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1535769750996_0010
18/09/01 12:16:14 INFO impl.YarnClientImpl: Submitted application application_1535769750996_0010
18/09/01 12:16:14 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1535769750996_0010/
18/09/01 12:16:14 INFO mapreduce.Job: Running job: job_1535769750996_0010
18/09/01 12:16:26 INFO mapreduce.Job: Job job_1535769750996_0010 running in uber mode : false
18/09/01 12:16:26 INFO mapreduce.Job:  map 0% reduce 0%
18/09/01 12:16:36 INFO mapreduce.Job:  map 100% reduce 0%
18/09/01 12:16:46 INFO mapreduce.Job:  map 100% reduce 100%
18/09/01 12:16:46 INFO mapreduce.Job: Job job_1535769750996_0010 completed successfully
18/09/01 12:16:46 INFO mapreduce.Job: Counters: 49
File System Counters
  FILE: Number of bytes read=96
  FILE: Number of bytes written=215319
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=345
  HDFS: Number of bytes written=37
  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)=6384
  Total time spent by all reduces in occupied slots (ms)=7857
  Total time spent by all map tasks (ms)=6384
  Total time spent by all reduce tasks (ms)=7857
  Total vcore-milliseconds taken by all map tasks=6384
  Total vcore-milliseconds taken by all reduce tasks=7857
  Total megabyte-milliseconds taken by all map tasks=6537216
  Total megabyte-milliseconds taken by all reduce tasks=8045568
Map-Reduce Framework
  Map input records=2
  Map output records=43
  Map output bytes=344
  Map output materialized bytes=96
  Input split bytes=122
  Combine input records=43
  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)=212
  CPU time spent (ms)=2120
  Physical memory (bytes) snapshot=308346880
  Virtual memory (bytes) snapshot=4118224896
  Total committed heap usage (bytes)=202379264
UNREGISTERED VERSION - Please support MobaXterm by subscribing to the professional edition here: https://mobaxterm.mobatek.net
```

```
192.168.43.242 (acadgild)
Terminal Sessions View X server Tools Games Settings Macros Help
Quick connect... 3 192.168.43.242 (acadgild) 4 192.168.43.242 (acadgild)
HDFS: Number of bytes read=345
HDFS: Number of bytes written=37
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)=6384
  Total time spent by all reduces in occupied slots (ms)=7857
  Total time spent by all map tasks (ms)=6384
  Total time spent by all reduce tasks (ms)=7857
  Total vcore-milliseconds taken by all map tasks=6384
  Total vcore-milliseconds taken by all reduce tasks=7857
  Total megabyte-milliseconds taken by all map tasks=6537216
  Total megabyte-milliseconds taken by all reduce tasks=8045568
Map-Reduce Framework
  Map input records=2
  Map output records=43
  Map output bytes=344
  Map output materialized bytes=96
  Input split bytes=122
  Combine input records=43
  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)=212
  CPU time spent (ms)=2120
  Physical memory (bytes) snapshot=308346880
  Virtual memory (bytes) snapshot=4118224896
  Total committed heap usage (bytes)=202379264
UNREGISTERED VERSION - Please support MobaXterm by subscribing to the professional edition here: https://mobaxterm.mobatek.net
```



The screenshot shows a terminal window with a dark background and white text. The text displays the output of a Hadoop MapReduce job. It includes statistics for the Map-Reduce Framework, such as input/output records, bytes, and splits. It also shows shuffle errors, file input/output counters, and job counters. The terminal is part of a MobaXterm application, with various icons and a search bar visible at the top and bottom.

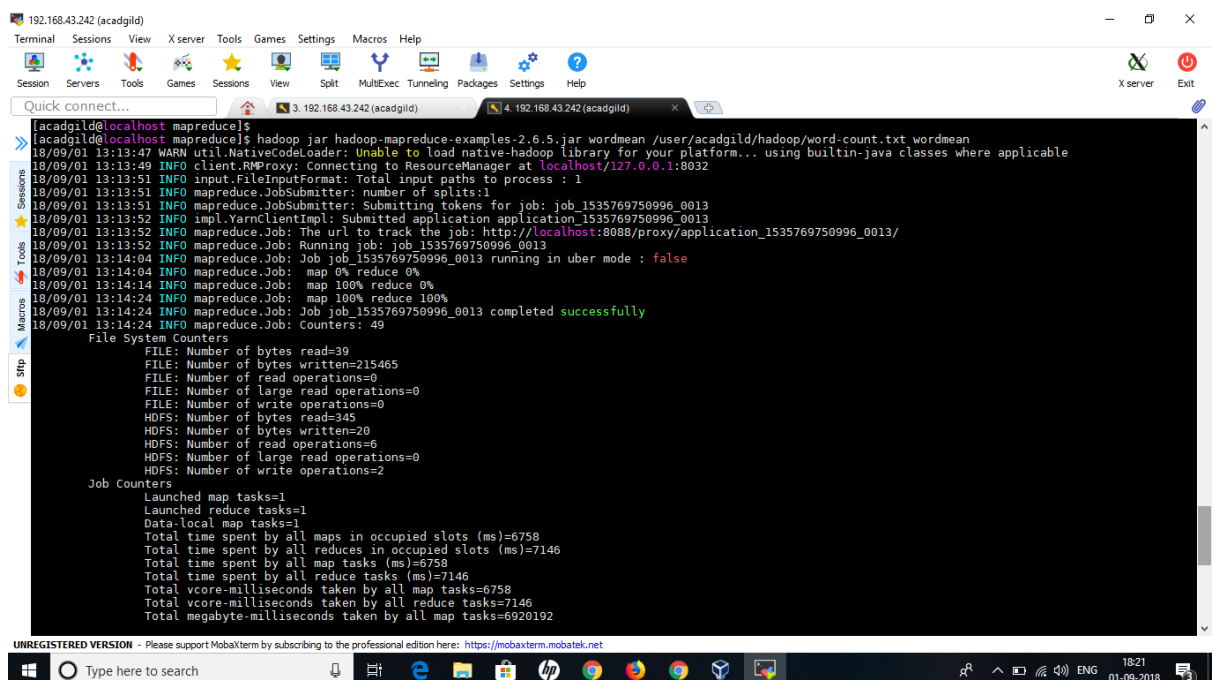
```
192.168.43.242 (acadgild)
Terminal Sessions View X server Tools Games Settings Macros Help
Session Servers Tools Games Sessions View Split MultiExec Tunneling Packages Settings Help
Quick connect... 3. 192.168.43.242 (acadgild) 4. 192.168.43.242 (acadgild)
Total vcore-milliseconds taken by all reduce tasks=7857
Total megabyte-milliseconds taken by all map tasks=6537216
Total megabyte-milliseconds taken by all reduce tasks=8045568
Map-Reduce Framework
  Map input records=2
  Map output records=43
  Map output bytes=344
  Map output materialized bytes=96
  Input split bytes=122
  Combine input records=43
  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)=212
  CPU time spent (ms)=2120
  Physical memory (bytes) snapshot=308346880
  Virtual memory (bytes) snapshot=4118224896
  Total committed heap usage (bytes)=202379264
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=223
File Output Format Counters
  Bytes Written=37
The median is: 4
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost mapreduce]$
```

2. **Word Mean:** A map/reduce program that counts the average length of the words in the input files.

Command: `hadoop jar hadoop-mapreduce-examples-2.6.5.jar wordmean /user/acadgild/hadoop/word-count.txt wordmean`

Explanation: We have used the map reduce jar file and mentioned the name of the program as wordmean then input file path as well as provide the output Path.

Command Execution:



The screenshot shows a terminal window with a dark background and white text. The text displays the execution of the wordmean command. It includes logs for the client connecting to the Resource Manager, the job submission, and the progress of the map and reduce tasks. The job is completed successfully. The terminal is part of a MobaXterm application, with various icons and a search bar visible at the top and bottom.

```
192.168.43.242 (acadgild)
Terminal Sessions View X server Tools Games Settings Macros Help
Session Servers Tools Games Sessions View Split MultiExec Tunneling Packages Settings Help
Quick connect... 3. 192.168.43.242 (acadgild) 4. 192.168.43.242 (acadgild)
[acadgild@localhost mapreduce]$
[acadgild@localhost mapreduce]$ hadoop jar hadoop-mapreduce-examples-2.6.5.jar wordmean /user/acadgild/hadoop/word-count.txt wordmean
18/09/01 13:13:47 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/09/01 13:13:49 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/09/01 13:13:51 INFO input.FileInputFormat: Total input paths to process : 1
18/09/01 13:13:51 INFO mapreduce.JobSubmitter: number of splits:1
18/09/01 13:13:51 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1535769750996_0013
18/09/01 13:13:52 INFO impl.YarnClientImpl: Submitted application application_1535769750996_0013
18/09/01 13:13:52 INFO mapreduce.Job: The url to track the job: http://localhost:8080/proxy/application_1535769750996_0013/
18/09/01 13:13:52 INFO mapreduce.Job: Running job: job_1535769750996_0013
18/09/01 13:14:04 INFO mapreduce.Job: Job job_1535769750996_0013 running in uber mode : false
18/09/01 13:14:04 INFO mapreduce.Job: map 0% reduce 0%
18/09/01 13:14:14 INFO mapreduce.Job: map 100% reduce 0%
18/09/01 13:14:24 INFO mapreduce.Job: map 100% reduce 100%
18/09/01 13:14:24 INFO mapreduce.Job: Job job_1535769750996_0013 completed successfully
18/09/01 13:14:24 INFO mapreduce.Job: Counters: 49
File System Counters
  FILE: Number of bytes read=39
  FILE: Number of bytes written=215465
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=345
  HDFS: Number of bytes written=20
  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)=6758
  Total time spent by all reduces in occupied slots (ms)=7146
  Total time spent by all map tasks (ms)=6758
  Total time spent by all reduce tasks (ms)=7146
  Total vcore-milliseconds taken by all map tasks=6758
  Total vcore-milliseconds taken by all reduce tasks=7146
  Total megabyte-milliseconds taken by all map tasks=6920192
```

```
192.168.43.242 (acadgild)
Terminal Sessions View X server Tools Games Settings Macros Help
Session Servers Tools Games Sessions View Split MultiExec Tunneling Packages Settings Help
Quick connect... 3. 192.168.43.242 (acadgild) 4. 192.168.43.242 (acadgild)

Total vcore-milliseconds taken by all reduce tasks=7146
Total megabyte-milliseconds taken by all map tasks=6020192
Total megabyte-milliseconds taken by all reduce tasks=7317504

Map-Reduce Framework
  Map input records=2
  Map output records=86
  Map output bytes=1247
  Map output materialized bytes=39
  Input split bytes=122
  Combine input records=86
  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)=166
  CPU time spent (ms)=1960
  Physical memory (bytes) snapshot=306954240
  Virtual memory (bytes) snapshot=4118224896
  Total committed heap usage (bytes)=202379264

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=223
File Output Format Counters
  Bytes Written=20
The mean is: 4.156946511627967
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost mapreduce]$
```

3. **Word Standard Deviation:** A map/reduce program that counts the standard deviation of the length of the words in the input files.

Command: `hadoop jar hadoop-mapreduce-examples-2.6.5.jar wordstandarddeviation /user/acadgild/hadoop/word-count.txt wordstandarddeviation`

Explanation: We have used the map reduce jar file and mentioned the name of the program as wordstandarddeviation then input file path as well as provide the output Path.

Command Execution:

```
192.168.43.242 (acadgild)
Terminal Sessions View X server Tools Games Settings Macros Help
Session Servers Tools Games Sessions View Split MultiExec Tunneling Packages Settings Help
Quick connect... 3. 192.168.43.242 (acadgild) 4. 192.168.43.242 (acadgild)

[acadgild@localhost mapreduce]$
[acadgild@localhost mapreduce]$ hadoop jar hadoop-mapreduce-examples-2.6.5.jar wordstandarddeviation /user/acadgild/hadoop/word-count.txt wordstandarddeviation
18/09/01 13:18:31 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/09/01 13:18:32 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/09/01 13:18:35 INFO input.FileInputFormat: Total input paths to process : 1
18/09/01 13:18:35 INFO mapreduce.JobSubmitter: number of splits:1
18/09/01 13:18:35 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1535769750996_0014
18/09/01 13:18:36 INFO impl.YarnClientImpl: Submitted application application_1535769750996_0014
18/09/01 13:18:36 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1535769750996_0014/
18/09/01 13:18:36 INFO mapreduce.Job: Running job: job_1535769750996_0014
18/09/01 13:18:54 INFO mapreduce.Job: Job job_1535769750996_0014 running in uber mode : false
18/09/01 13:18:54 INFO mapreduce.Job: map 0% reduce 0%
18/09/01 13:19:03 INFO mapreduce.Job: map 100% reduce 0%
18/09/01 13:19:13 INFO mapreduce.Job: map 100% reduce 100%
18/09/01 13:19:14 INFO mapreduce.Job: Job job_1535769750996_0014 completed successfully
18/09/01 13:19:15 INFO mapreduce.Job: Counters: 49

File System Counters
  FILE: Number of bytes read=56
  FILE: Number of bytes written=215685
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=345
  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)=7029
  Total time spent by all reduces in occupied slots (ms)=8062
  Total time spent by all map tasks (ms)=7029
  Total time spent by all reduce tasks (ms)=8062
  Total vcore-milliseconds taken by all map tasks=7029
  Total vcore-milliseconds taken by all reduce tasks=8062
  Total megabyte-milliseconds taken by all map tasks=7197696
```

192.168.43.242 (acadgild)

Terminal Sessions View X server Tools Games Settings Macros Help

Session Servers Tools Games Sessions View Split MultiExec Tunneling Packages Settings Help

Quick connect...

3. 192.168.43.242 (acadgild) 4. 192.168.43.242 (acadgild)

```
Total vcore-milliseconds taken by all reduce tasks=8862
Total megabyte-milliseconds taken by all map tasks=7197696
Total megabyte-milliseconds taken by all reduce tasks=8255488

Map-Reduce Framework
  Map input records=2
  Map output records=129
  Map output bytes=1892
  Map output materialized bytes=56
  Input split bytes=122
  Combine input records=129
  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)=216
  CPU time spent (ms)=2050
  Physical memory (bytes) snapshot=308989952
  Virtual memory (bytes) snapshot=4118216794
  Total committed heap usage (bytes)=282379264

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=223
File Output Format Counters
  Bytes Written=31
The standard deviation is: 1.932052938101063
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost mapreduce]$
```

UNREGISTERED VERSION - Please support MobaXterm by subscribing to the professional edition here: <https://mobaxterm.mobatek.net>

Type here to search

18-25 01-09-2018