Task 1

pwd - It gives us the absolute path, which means the path that starts from the root.

```
[cloudera@quickstart Desktop]$ pwd
/home/cloudera/Desktop
```

vi -vi is basically a text editor

```
[cloudera@quickstart Desktop]$ vi tawanda.txt
[cloudera@quickstart Desktop]$ cat tawanda.txt
tsslsls
lls
lsls
jjj
nnn
sss
fity
nhum
kkk
gggu
suss

[cloudera@quickstart Desktop]$
```

touch — The touch command is used to create a file.

ls — Use the "ls" command to know what files are in the directory you are in

echo — The "echo" command helps us move some data, usually text into a file.

cat — Use the cat command to display the contents of a file.

```
[cloudera@quickstart Desktop]$ ls
collecttimetopology.txt  Eclipse.desktop  Enterprise.desktop  Express.desktop  Kerberos.desktop  LTE_Counter_Report_Daily.xlsx  Parcels.desktop  test_file
[cloudera@quickstart Desktop]$ touch cloudera.txt
[cloudera@quickstart Desktop]$ ls
cloudera.txt  collecttimetopology.txt  Eclipse.desktop  Enterprise.desktop  Express.desktop  Kerberos.desktop  LTE_Counter_Report_Daily.xlsx  Parcels.desktop  test_file
[cloudera@quickstart Desktop]$ echo My name is Tawanda >> cloudera.txt
[cloudera@quickstart Desktop]$ cat cloudera.txt
My name is Tawanda
[cloudera@quickstart Desktop]$
```

cd — Use the "cd" command to go to a directory.

mkdir - Use the mkdir command when you need to create a folder or a directory.

rm - Use the rm command to delete files and directories.

mv - move - allows a user to move a file to another folder or directory

cp — Use the cp command to copy files through the command line

```
[cloudera@quickstart Desktop]$ mkdir Tawanda
[cloudera@quickstart Desktop]$ ls
cloudera.txt  collecttimetopology.txt  Eclipse.desktop  Enterprise.desktop  Express.desktop  Kerberos.desktop  LTE_Counter_Report_Daily.xlsx  Parcels.desktop  Tawanda  test_file
[cloudera@quickstart Desktop]$ cd Tawanda
[cloudera@quickstart Tawanda]$ cd ..
[cloudera@quickstart Desktop]$ ls
cloudera.txt  collecttimetopology.txt  Eclipse.desktop  Enterprise.desktop  Express.desktop  Kerberos.desktop  LTE_Counter_Report_Daily.xlsx  Parcels.desktop  Tawanda  test_file
[cloudera@quickstart Desktop]$ mv cloudera.txt /home/cloudera/Desktop/Tawanda/
[cloudera@quickstart Desktop]$ cd Tawanda
[cloudera@quickstart Tawanda]$ ls
cloudera.txt
[cloudera@quickstart Tawanda]$ cp cloudera.txt /home/cloudera/Desktop/
[cloudera@quickstart Tawanda]$ cd ..
[cloudera@quickstart Desktop]$ ls
cloudera.txt  collecttimetopology.txt  Eclipse.desktop  Enterprise.desktop  Express.desktop  Kerberos.desktop  LTE_Counter_Report_Daily.xlsx  Parcels.desktop  Tawanda  test_file
[cloudera@quickstart Desktop]$ rm cloudera.txt
rm: remove regular file `cloudera.txt'?
[cloudera@quickstart Desktop]$ ls
cloudera.txt  collecttimetopology.txt  Eclipse.desktop  Enterprise.desktop  Express.desktop  Kerberos.desktop  LTE_Counter_Report_Daily.xlsx  Parcels.desktop  Tawanda  test_file
[cloudera@quickstart Desktop]$
```

date - date command can be used to print (or even set) the system date and time.

cal - The cal command display a calendar in the output.

mv - mv command lets you either move a file from one directory to another, or rename it.

which - which command basically lets you locate a command - the file and the path of the file that gets executed

who - who command shows who is logged on.

```
[cloudera@quickstart Desktop]$ date
Tue Mar 19 00:59:49 PDT 2019
[cloudera@quickstart Desktop]$ cal
     March 2019
Su Mo Tu We Th Fr Sa
                1  2
 3  4  5  6  7  8  9
10 11 12 13 14 15 16
17 18 19 20 21 22 23
24 25 26 27 28 29 30
31
[cloudera@quickstart Desktop]$ which date
/bin/date
[cloudera@quickstart Desktop]$ who
cloudera tty1         2019-03-18 23:58 (:0)
cloudera pts/0        2019-03-19 00:02 (:0.0)
[cloudera@quickstart Desktop]$
```

Task 2

cloudera@quickstart hadoop-mapreduce]$ hadoop jar /usr/lib/hadoop-mapreduce/hadoop-mapreduce-examples-2.6.0-cdh5.13.0.jar wordmean /input /outputmean

19/03/19 23:52:16 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032

19/03/19 23:52:17 INFO input.FileInputFormat: Total input paths to process : 1

19/03/19 23:52:17 INFO mapreduce.JobSubmitter: number of splits:1

19/03/19 23:52:17 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1552978605877_0002

19/03/19 23:52:18 INFO impl.YarnClientImpl: Submitted application application_1552978605877_0002

19/03/19 23:52:18 INFO mapreduce.Job: The url to track the job:
http://quickstart.cloudera:8088/proxy/application_1552978605877_0002/

19/03/19 23:52:18 INFO mapreduce.Job: Running job: job_1552978605877_0002

19/03/19 23:52:26 INFO mapreduce.Job: Job job_1552978605877_0002 running in uber mode : false

19/03/19 23:52:26 INFO mapreduce.Job: map 0% reduce 0%

19/03/19 23:52:35 INFO mapreduce.Job: map 100% reduce 0%

19/03/19 23:52:43 INFO mapreduce.Job: map 100% reduce 100%

19/03/19 23:52:44 INFO mapreduce.Job: Job job_1552978605877_0002 completed successfully

19/03/19 23:52:44 INFO mapreduce.Job: Counters: 49

File System Counters

FILE: Number of bytes read=39

FILE: Number of bytes written=287259

FILE: Number of read operations=0

FILE: Number of large read operations=0

FILE: Number of write operations=0

HDFS: Number of bytes read=4014431

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)=5805

Total time spent by all reduces in occupied slots (ms)=5451

Total time spent by all map tasks (ms)=5805

Total time spent by all reduce tasks (ms)=5451

Total vcore-milliseconds taken by all map tasks=5805

Total vcore-milliseconds taken by all reduce tasks=5451

Total megabyte-milliseconds taken by all map tasks=5944320

Total megabyte-milliseconds taken by all reduce tasks=5581824

Map-Reduce Framework

Map input records=129494

Map output records=776962

Map output bytes=11265949

Map output materialized bytes=39

Input split bytes=126

Combine input records=776962

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)=115

CPU time spent (ms)=4300

Physical memory (bytes) snapshot=479567872

Virtual memory (bytes) snapshot=3151101952

Total committed heap usage (bytes)=436731904

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=4014305

File Output Format Counters

Bytes Written=28

The mean is: 9.000002574128464

[cloudera@quickstart hadoop-mapreduce]$ hadoop jar /usr/lib/hadoop-mapreduce/hadoop-mapreduce-examples-2.6.0-cdh5.13.0.jar wordmedian /input /outputmedian

19/03/19 23:53:38 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032

19/03/19 23:53:39 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.

19/03/19 23:53:39 INFO input.FileInputFormat: Total input paths to process : 1

19/03/19 23:53:39 INFO mapreduce.JobSubmitter: number of splits:1

19/03/19 23:53:40 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1552978605877_0003

19/03/19 23:53:40 INFO impl.YarnClientImpl: Submitted application application_1552978605877_0003

19/03/19 23:53:40 INFO mapreduce.Job: The url to track the job:
http://quickstart.cloudera:8088/proxy/application_1552978605877_0003/

19/03/19 23:53:40 INFO mapreduce.Job: Running job: job_1552978605877_0003

19/03/19 23:53:48 INFO mapreduce.Job: Job job_1552978605877_0003 running in uber mode : false

19/03/19 23:53:48 INFO mapreduce.Job: map 0% reduce 0%

19/03/19 23:53:57 INFO mapreduce.Job: map 100% reduce 0%

19/03/19 23:54:05 INFO mapreduce.Job: map 100% reduce 100%

19/03/19 23:54:05 INFO mapreduce.Job: Job job_1552978605877_0003 completed successfully

19/03/19 23:54:06 INFO mapreduce.Job: Counters: 49

File System Counters

FILE: Number of bytes read=56

FILE: Number of bytes written=286985

FILE: Number of read operations=0

FILE: Number of large read operations=0

FILE: Number of write operations=0

HDFS: Number of bytes read=4014431

HDFS: Number of bytes written=38

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)=5921

Total time spent by all reduces in occupied slots (ms)=4908

Total time spent by all map tasks (ms)=5921

Total time spent by all reduce tasks (ms)=4908

Total vcore-milliseconds taken by all map tasks=5921

Total vcore-milliseconds taken by all reduce tasks=4908

Total megabyte-milliseconds taken by all map tasks=6063104

Total megabyte-milliseconds taken by all reduce tasks=5025792

Map-Reduce Framework

Map input records=129494

Map output records=388481

Map output bytes=3107848

Map output materialized bytes=56

Input split bytes=126

Combine input records=388481

Combine output records=5

Reduce input groups=5

Reduce shuffle bytes=56

Reduce input records=5

Reduce output records=5

Spilled Records=10

Shuffled Maps =1

Failed Shuffles=0

Merged Map outputs=1

GC time elapsed (ms)=67

CPU time spent (ms)=4430

Physical memory (bytes) snapshot=522555392

Virtual memory (bytes) snapshot=3154231296

Total committed heap usage (bytes)=435159040

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=4014305

File Output Format Counters

Bytes Written=38

The median is: 10

[cloudera@quickstart hadoop-mapreduce]$ hadoop jar /usr/lib/hadoop-mapreduce/hadoop-mapreduce-examples-2.6.0-cdh5.13.0.jar wordstandarddeviation /input /outputstd

19/03/20 00:14:38 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032

19/03/20 00:14:39 INFO input.FileInputFormat: Total input paths to process : 1

19/03/20 00:14:39 INFO mapreduce.JobSubmitter: number of splits:1

19/03/20 00:14:39 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1552978605877_0004

19/03/20 00:14:40 INFO impl.YarnClientImpl: Submitted application application_1552978605877_0004

19/03/20 00:14:40 INFO mapreduce.Job: The url to track the job:
http://quickstart.cloudera:8088/proxy/application_1552978605877_0004/

19/03/20 00:14:40 INFO mapreduce.Job: Running job: job_1552978605877_0004

19/03/20 00:14:49 INFO mapreduce.Job: Job job_1552978605877_0004 running in uber mode : false

19/03/20 00:14:49 INFO mapreduce.Job: map 0% reduce 0%

19/03/20 00:14:57 INFO mapreduce.Job: map 100% reduce 0%

19/03/20 00:15:06 INFO mapreduce.Job: map 100% reduce 100%

19/03/20 00:15:06 INFO mapreduce.Job: Job job_1552978605877_0004 completed successfully

19/03/20 00:15:06 INFO mapreduce.Job: Counters: 49

File System Counters

FILE: Number of bytes read=56

FILE: Number of bytes written=287451

FILE: Number of read operations=0

FILE: Number of large read operations=0

FILE: Number of write operations=0

HDFS: Number of bytes read=4014431

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)=6419

Total time spent by all reduces in occupied slots (ms)=5099

Total time spent by all map tasks (ms)=6419

Total time spent by all reduce tasks (ms)=5099

Total vcore-milliseconds taken by all map tasks=6419

Total vcore-milliseconds taken by all reduce tasks=5099

Total megabyte-milliseconds taken by all map tasks=6573056

Total megabyte-milliseconds taken by all reduce tasks=5221376

Map-Reduce Framework

Map input records=129494

Map output records=1165443

Map output bytes=17093164

Map output materialized bytes=56

Input split bytes=126

Combine input records=1165443

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)=74

CPU time spent (ms)=5000

Physical memory (bytes) snapshot=527618048

Virtual memory (bytes) snapshot=3145359360

Total committed heap usage (bytes)=433586176

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=4014305

File Output Format Counters

Bytes Written=44

The standard deviation is: 2.943914896710781

[cloudera@quickstart hadoop-mapreduce]$