

Task1

1. Pwd - reports the full path to the current directory

```
Last login: Tue Feb 19 00:06:13 2019 from 10.0.2.2
[cloudera@quickstart ~]$
[cloudera@quickstart ~]$
[cloudera@quickstart ~]$
[cloudera@quickstart ~]$ pwd
/home/cloudera
[cloudera@quickstart ~]$
```

2. Vi – writes in a file

```
[root@quickstart cloudera]# vi test11

Hi my name is leo
~
~
~
~
~
```

3. Touch- creates a file

```
[root@quickstart cloudera]# touch test
[root@quickstart cloudera]#
```

4. Mkdir – creates a directory

```
[root@quickstart cloudera]# mkdir /test1
[root@quickstart cloudera]#
```

5. Rm – deletes file

```
[root@quickstart cloudera]# rm test
rm: remove regular empty file `test'? y
[root@quickstart cloudera]#
```

6. Ls- list all files and sub directories in a directory

```
[root@quickstart cloudera]# ls
cloudera-manager  complex  Downloads  empl.csv
cm_api.py         Desktop  eclipse    enterprise-deployment.json
company.java      Documents emp_data.csv express-deployment.json
[root@quickstart cloudera]#
```

7. Echo- writes its arguments to standard output

```
[root@quickstart cloudera]# echo test2
test2
[root@quickstart cloudera]#
```

8. Cat – reads through a file

```
[root@quickstart cloudera]# cat test2
ddjsdhjshd
[root@quickstart cloudera]#
```

9. Who - used to find out the following information :
1. Time of last system boot
 2. Current run level of the system
 3. List of logged in users and more.

```
[root@quickstart cloudera]# who
cloudera tty1      2019-02-25 05:53 (:0)
cloudera pts/0     2019-02-25 05:54 (10.0.2.2)
[root@quickstart cloudera]#
```

10. Cd – change directory

```
[root@quickstart cloudera]# cd Downloads/
[root@quickstart Downloads]#
```

11. Date – show todays date and time

```
[root@quickstart Downloads]# date
Mon Feb 25 06:33:06 PST 2019
[root@quickstart Downloads]#
```

12. Cal- shows this month calendar

```
[root@quickstart Downloads]# cal
      February 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

[root@quickstart Downloads]#
```

13. Mv – moves files to a desired location

```
[root@quickstart cloudera]# mv test11 Videos
[root@quickstart cloudera]#
```

14. Cp – copies files to desired location

```
[root@quickstart cloudera]# cp test2 Videos
[root@quickstart cloudera]#
```

15. Which - returns the pathnames of the files (or links) which would be executed in the current environment

```
[root@quickstart cloudera]# which sh
/bin/sh
[root@quickstart cloudera]#
```

Task 2

Wordmedian

```
[ec2-54-93-100-100@localhost ~]$ hadoop jar hadoop-mapreduce-examples-2.6.5.jar wordmedian test.log word.out
19/03/27 16:16:19 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
19/03/27 16:16:20 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
19/03/27 16:16:20 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
19/03/27 16:16:21 INFO input.FileInputFormat: Total input paths to process : 1
19/03/27 16:16:21 INFO mapreduce.JobSubmitter: number of splits:1
19/03/27 16:16:21 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1553681483018_0001
19/03/27 16:16:22 INFO impl.YarnClientImpl: Submitted application application_1553681483018_0001
19/03/27 16:16:22 INFO mapreduce.Job: The url to track the job: http://localhost:8080/proxy/application_1553681483018_0001/
19/03/27 16:16:22 INFO mapreduce.Job: Running job: job_1553681483018_0001
19/03/27 16:16:31 INFO mapreduce.Job: Job job_1553681483018_0001 running in uber mode : false
19/03/27 16:16:31 INFO mapreduce.Job:  map 0% reduce 0%
19/03/27 16:16:36 INFO mapreduce.Job:  map 100% reduce 0%
19/03/27 16:16:42 INFO mapreduce.Job:  map 100% reduce 100%
19/03/27 16:16:42 INFO mapreduce.Job: Job job_1553681483018_0001 completed successfully
19/03/27 16:16:43 INFO mapreduce.Job: Counters: 49

  File System Counters
    FILE: Number of bytes read=246
    FILE: Number of bytes written=215509
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=7956
    HDFS: Number of bytes written=122
    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)=3119
    Total time spent by all reduces in occupied slots (ms)=3405
    Total time spent by all map tasks (ms)=3119
    Total time spent by all reduce tasks (ms)=3405
    Total vcore-milliseconds taken by all map tasks=3119
    Total vcore-milliseconds taken by all reduce tasks=3405
    Total megabyte-milliseconds taken by all map tasks=3193856
    Total megabyte-milliseconds taken by all reduce tasks=3486720
```

```

Total megabyte-milliseconds taken by all Reduce tasks=3160720
Map-Reduce Framework
  Map input records=312
  Map output records=678
  Map output bytes=5424
  Map output materialized bytes=246
  Input split bytes=100
  Combine input records=678
  Combine output records=24
  Reduce input groups=24
  Reduce shuffle bytes=246
  Reduce input records=24
  Reduce output records=24
  Spilled Records=48
  Shuffled Maps =1
  Failed Shuffles=0
  Merged Map outputs=1
  GC time elapsed (ms)=161
  CPU time spent (ms)=1980
  Physical memory (bytes) snapshot=426643456
  Virtual memory (bytes) snapshot=4183851008
  Total committed heap usage (bytes)=322437120
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=7847
File Output Format Counters
  Bytes Written=122
The median is: 3

```

Wordmean

```

[acaddi@localhost mapreduce]$ hadoop jar hadoop-mapreduce-examples-2.6.5.jar wordmean test.log wordmean.out
19/03/27 16:19:19 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
19/03/27 16:19:20 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
19/03/27 16:19:21 INFO input.FileInputFormat: Total input paths to process : 1
19/03/27 16:19:21 INFO mapreduce.JobSubmitter: number of splits:1
19/03/27 16:19:21 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1553681403018_0002
19/03/27 16:19:21 INFO impl.YarnClientImpl: Submitted application application_1553681403018_0002
19/03/27 16:19:21 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1553681403018_0002/
19/03/27 16:19:21 INFO mapreduce.Job: Running job: job_1553681403018_0002
19/03/27 16:19:27 INFO mapreduce.Job: Job job_1553681403018_0002 running in uber mode : false
19/03/27 16:19:27 INFO mapreduce.Job:  map 0% reduce 0%
19/03/27 16:19:32 INFO mapreduce.Job:  map 100% reduce 0%
19/03/27 16:19:39 INFO mapreduce.Job:  map 100% reduce 100%
19/03/27 16:19:40 INFO mapreduce.Job: Job job_1553681403018_0002 completed successfully
19/03/27 16:19:40 INFO mapreduce.Job: Counters: 49
File System Counters
  FILE: Number of bytes read=39
  FILE: Number of bytes written=215447
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=7956
  HDFS: Number of bytes written=22
  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)=2859
  Total time spent by all reduces in occupied slots (ms)=3355
  Total time spent by all map tasks (ms)=2859
  Total time spent by all reduce tasks (ms)=3355
  Total vcore-milliseconds taken by all map tasks=2859
  Total vcore-milliseconds taken by all reduce tasks=3355
  Total megabyte-milliseconds taken by all map tasks=2927616

```

```

Map-Reduce Framework
  Map input records=312
  Map output records=1356
  Map output bytes=19662
  Map output materialized bytes=39
  Input split bytes=109
  Combine input records=1356
  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)=159
  CPU time spent (ms)=1850
  Physical memory (bytes) snapshot=419135488
  Virtual memory (bytes) snapshot=4183732224
  Total committed heap usage (bytes)=305135616
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=7847
File Output Format Counters
  Bytes Written=22
The mean is: 9.03834808259587

```

WordStandard deviation

```
[acadgild@localhost mapreduce]$ hadoop jar hadoop-mapreduce-examples-2.6.5.jar wordstandarddeviation test.log wordSDDev.out
19/03/27 16:21:38 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
19/03/27 16:21:38 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
19/03/27 16:21:39 INFO input.FileInputFormat: Total input paths to process : 1
19/03/27 16:21:39 INFO mapreduce.JobSubmitter: number of splits:1
19/03/27 16:21:39 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1553681403018_0003
19/03/27 16:21:40 INFO impl.YarnClientImpl: Submitted application application_1553681403018_0003
19/03/27 16:21:40 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1553681403018_0003/
19/03/27 16:21:40 INFO mapreduce.Job: Running job: job_1553681403018_0003
19/03/27 16:21:46 INFO mapreduce.Job: Job job_1553681403018_0003 running in uber mode : false
19/03/27 16:21:46 INFO mapreduce.Job: map 0% reduce 0%
19/03/27 16:21:51 INFO mapreduce.Job: map 100% reduce 0%
19/03/27 16:21:57 INFO mapreduce.Job: map 100% reduce 100%
19/03/27 16:21:58 INFO mapreduce.Job: Job job_1553681403018_0003 completed successfully
19/03/27 16:21:59 INFO mapreduce.Job: Counters: 49
  File System Counters
    FILE: Number of bytes read=56
    FILE: Number of bytes written=215643
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=7956
    HDFS: Number of bytes written=36
    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)=2979
    Total time spent by all reduces in occupied slots (ms)=3104
    Total time spent by all map tasks (ms)=2979
    Total time spent by all reduce tasks (ms)=3104
    Total vcore-milliseconds taken by all map tasks=2979
    Total vcore-milliseconds taken by all reduce tasks=3104
    Total megabyte-milliseconds taken by all map tasks=3050496
    Total megabyte-milliseconds taken by all reduce tasks=3178496
```

```
    Total megabyte-milliseconds taken by all reduce tasks=3178496
Map-Reduce Framework
  Map input records=312
  Map output records=2034
  Map output bytes=29832
  Map output materialized bytes=56
  Input split bytes=109
  Combine input records=2034
  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)=134
  CPU time spent (ms)=1920
  Physical memory (bytes) snapshot=419663872
  Virtual memory (bytes) snapshot=4184227840
  Total committed heap usage (bytes)=302514176
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=7847
File Output Format Counters
  Bytes Written=36
The standard deviation is: 17.097197537691677
```