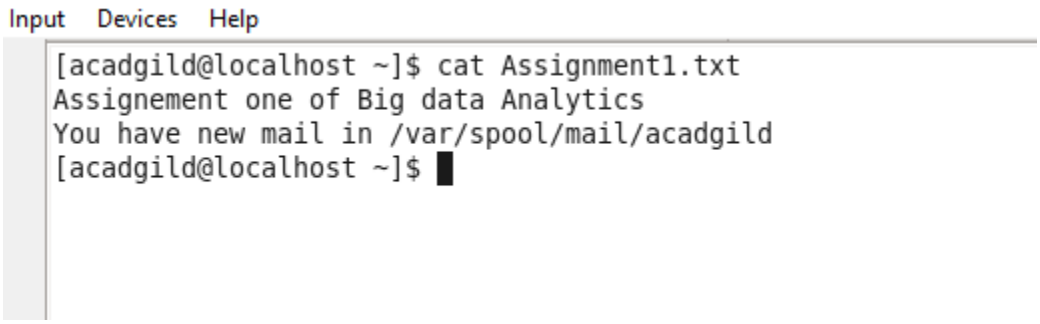
Assignment 3

Task1: Execute WordMean program using the jar file present in VM

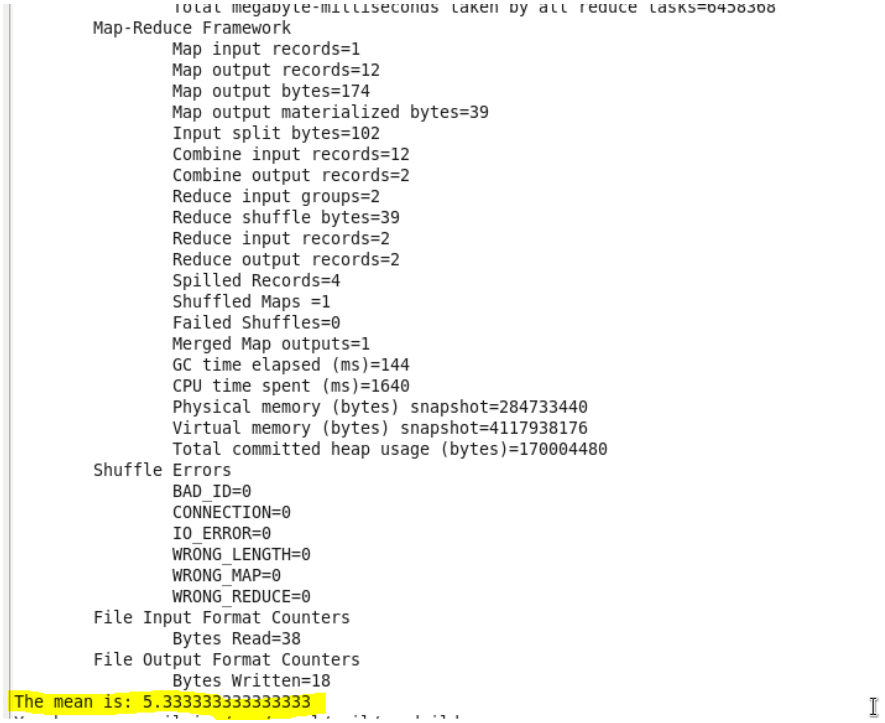
The file used – Assignment1. Txt

Here is a screenshot of the content saved in the file for which the wordmean is to be found-



The command used: "Hadoop jar /home/acadgild/install/Hadoop/Hadoop-2.6.5/share/Hadoop/mapreduce/Hadoop-mapreduce-examples-2.6.5.jar word mean /Assignment1.txt / newoutput"

Here is the output obtained for finding wordmean-



Task2: Execute WordMedian program using the jar file present in VM

Command used: "Hadoop jar /home/acadgild/install/Hadoop/Hadoop-2.6.5/share/Hadoop/mapreduce/Hadoop-mapreduce-examples-2.6.5.jar word median /Assignment1.txt / output1"

Here is a screenshot of the content saved in the file for which the wordmedian is to be found-

```
                Total megabyte-milliseconds taken by all map tasks=4784128
                Total megabyte-milliseconds taken by all reduce tasks=5399552
        Map-Reduce Framework
                Map input records=1
                Map output records=6
                Map output bytes=48
                Map output materialized bytes=56
                Input split bytes=102
                Combine input records=6
                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)=143
                CPU time spent (ms)=1400
                Physical memory (bytes) snapshot=295661568
                Virtual memory (bytes) snapshot=4118228992
                Total committed heap usage (bytes)=170004480
        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=38
        File Output Format Counters
                Bytes Written=21
The median is: 3
```

Task3: Execute WordStandardDeviation program using the jar file present in VM

Command used: "Hadoop jar /home/acadgild/install/Hadoop/Hadoop-2.6.5/share/Hadoop/mapreduce/Hadoop-mapreduce-examples-2.6.5.jar word standard deviation /Assignment1.txt / output2"

Here is a screenshot of the content saved in the file for which the wordmedian is to be found-

```
        Map-Reduce Framework
                Map input records=1
                Map output records=18
                Map output bytes=264
                Map output materialized bytes=56
                Input split bytes=102
                Combine input records=18
                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)=133
                CPU time spent (ms)=1340
                Physical memory (bytes) snapshot=298065920
                Virtual memory (bytes) snapshot=4118216704
                Total committed heap usage (bytes)=170004480
        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=38
        File Output Format Counters
                Bytes Written=29
The standard deviation is: 3.39934634239519
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$
```