Problem Statement

Execute WordMedian , WordMean , WordStandardDeviation programs using 1hadoop-mapreduce-examples-2.9.0.jar file present in your AcadGild VM.

- Create a text File containing N Number of Words in local File System.

  Command: Vi wordcount.txt

- Put the file to hdfs File System
  Command:
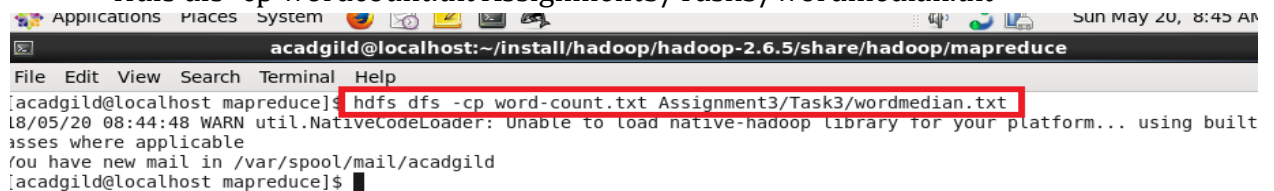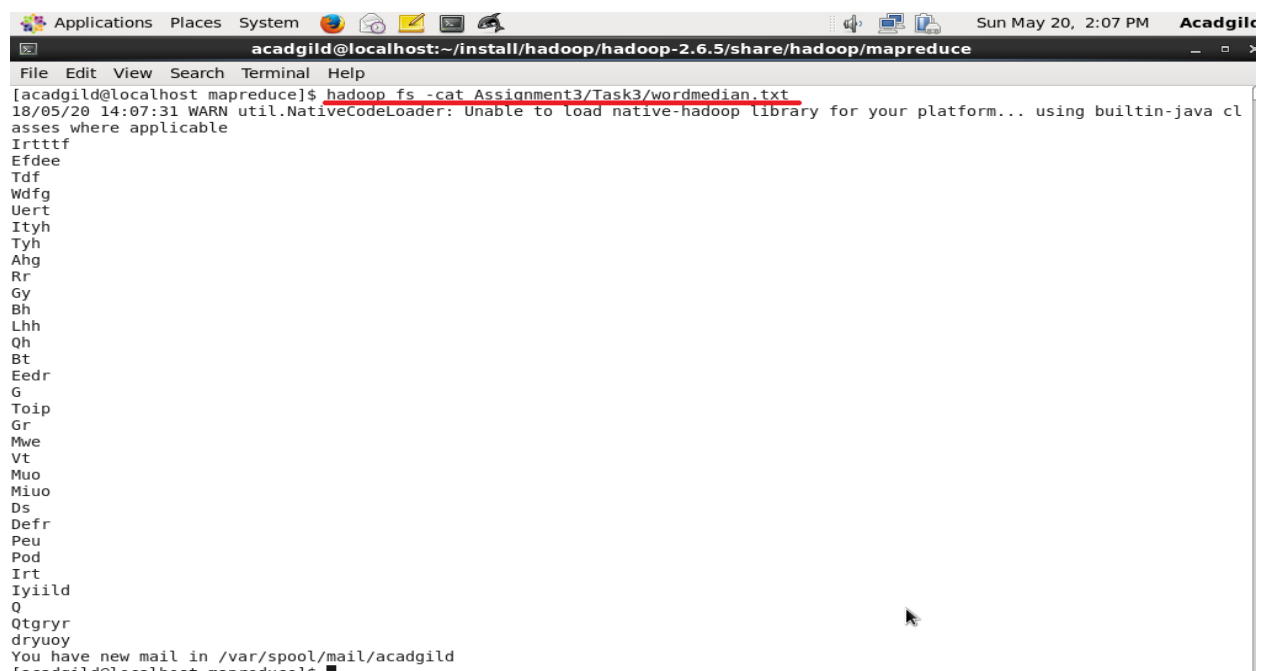 Hdfs dfs -put wordcount.txt /user/acadgiid/wordcount.txt

1. **Word Median**
   Step1:
   Copy the file wordcount.txt to wordmedian.txt

   Command:

   Hdfs dfs -cp wordcount.txt Assignment3/Task3/wordmedian.txt

<u>Step2:</u>

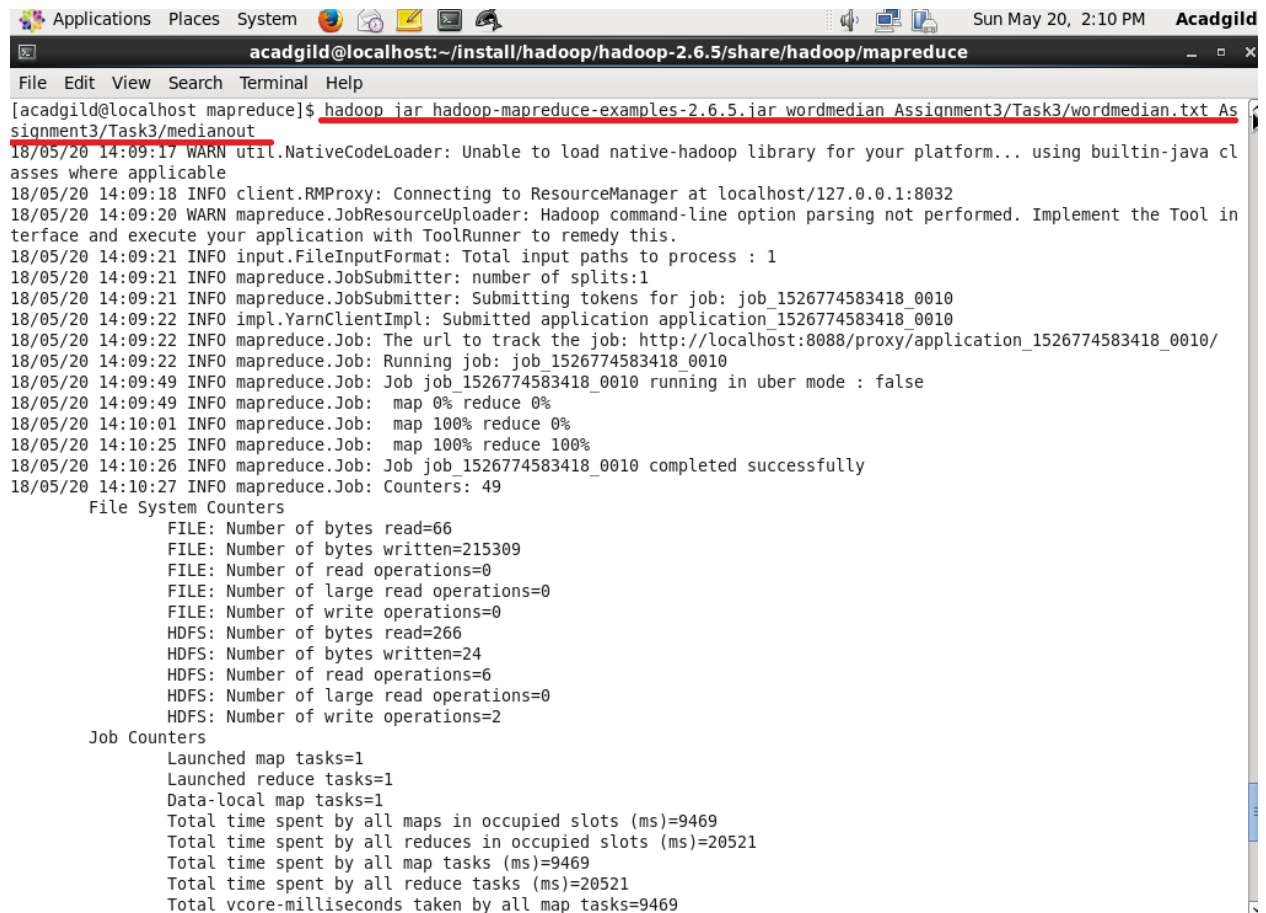Execute the jar file for word median.

Syntax:

Hadoop jar <jarfilename> <mainclass name> <inputfile path> <outputFilePath>

Command:

Hadoop jar Hadoop-mapreduce-examples-2.5.6.jar wordmedian Assignment3/Task3/wordmedian.txt Assignment3/Task3/medianout

```
              Total vcore-milliseconds taken by all map tasks=9469
              Total vcore-milliseconds taken by all reduce tasks=20521
              Total megabyte-milliseconds taken by all map tasks=9696256
              Total megabyte-milliseconds taken by all reduce tasks=21013504
      Map-Reduce Framework
              Map input records=31
              Map output records=31
              Map output bytes=248
              Map output materialized bytes=66
              Input split bytes=133
              Combine input records=31
              Combine output records=6
              Reduce input groups=6
              Reduce shuffle bytes=66
              Reduce input records=6
              Reduce output records=6
              Spilled Records=12
              Shuffled Maps =1
              Failed Shuffles=0
              Merged Map outputs=1
              GC time elapsed (ms)=458
              CPU time spent (ms)=4340
              Physical memory (bytes) snapshot=297959424
              Virtual memory (bytes) snapshot=4118220800
              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=133
      File Output Format Counters
              Bytes Written=24
The median is: 3
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost mapreduce]$ ▮
```

Step 3:
Display the contents of output File

```
[acadgild@localhost mapreduce]$ hadoop fs -ls Assignment3/Task3/medianout
18/05/20 14:12:30 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
Found 2 items
-rw-r--r--   1 acadgild supergroup          0 2018-05-20 14:10 Assignment3/Task3/medianout/_SUCCESS
-rw-r--r--   1 acadgild supergroup         24 2018-05-20 14:10 Assignment3/Task3/medianout/part-r-00000
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost mapreduce]$ hadoop fs -cat Assignment3/Task3/medianout/part-r-00000
18/05/20 14:13:06 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
1       2
2       8
3       9
4       7
5       1
6       4
[acadgild@localhost mapreduce]$ ▮
```
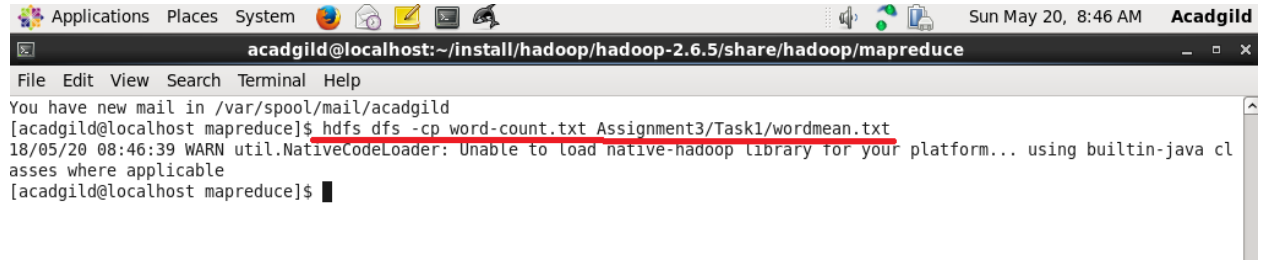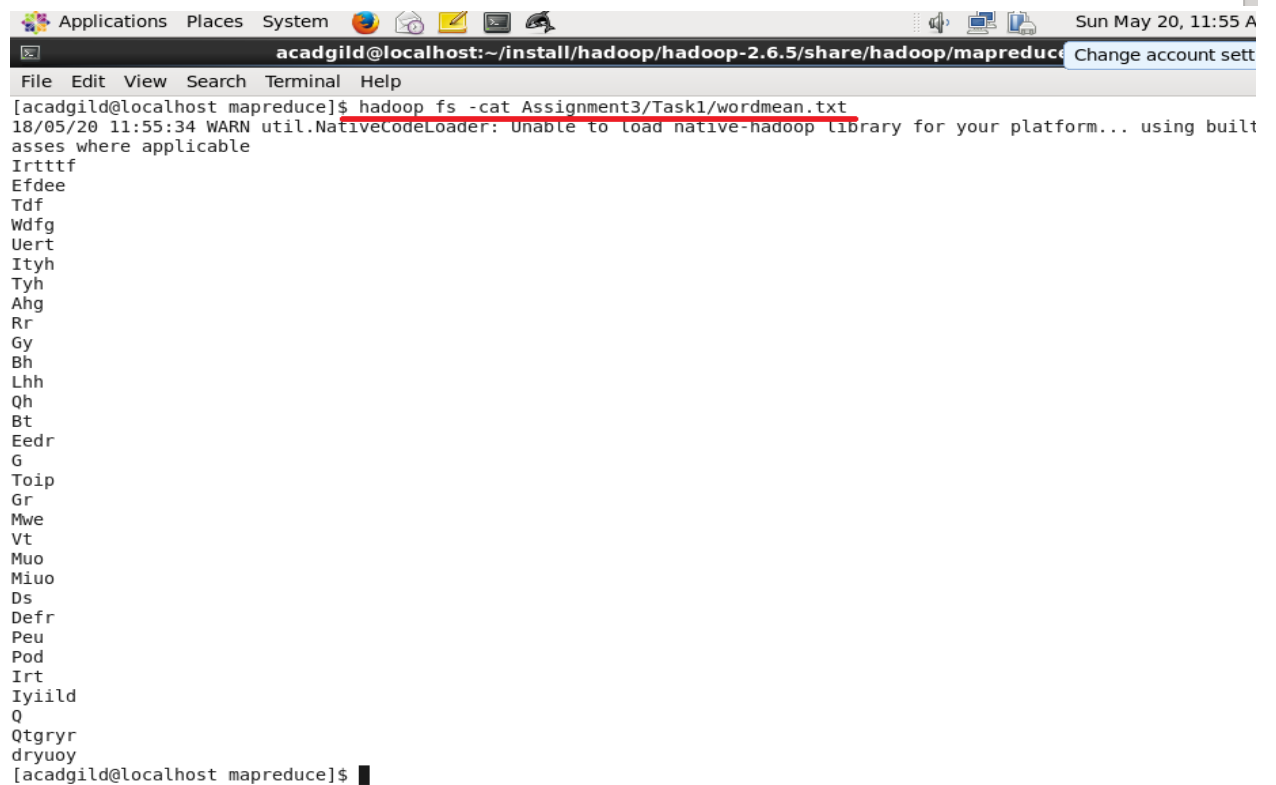
2. Word Mean

<u>Step1:</u>

Copy the file wordcount.txt to wordmean.txt

Command:

Hdfs dfs -cp wordcount.txt Assignment3/Task1/wordmean.txt

```
Applications  Places  System                              Sun May 20, 8:46 AM  Acadgild

           acadgild@localhost:~/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce       _  □  ×

File  Edit  View  Search  Terminal  Help
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost mapreduce]$ hdfs dfs -cp word-count.txt Assignment3/Task1/wordmean.txt
18/05/20 08:46:39 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
[acadgild@localhost mapreduce]$ █
```

```
Applications  Places  System                              Sun May 20, 11:55 A

           acadgild@localhost:~/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce   Change account sett

File  Edit  View  Search  Terminal  Help
[acadgild@localhost mapreduce]$ hadoop fs -cat Assignment3/Task1/wordmean.txt
18/05/20 11:55:34 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using built
asses where applicable
Irtttf
Efdee
Tdf
Wdfg
Uert
Ityh
Tyh
Ahg
Rr
Gy
Bh
Lhh
Qh
Bt
Eedr
G
Toip
Gr
Mwe
Vt
Muo
Miuo
Ds
Defr
Peu
Pod
Irt
Iyiild
Q
Qtgryr
dryuoy
[acadgild@localhost mapreduce]$ █
```

Step2:

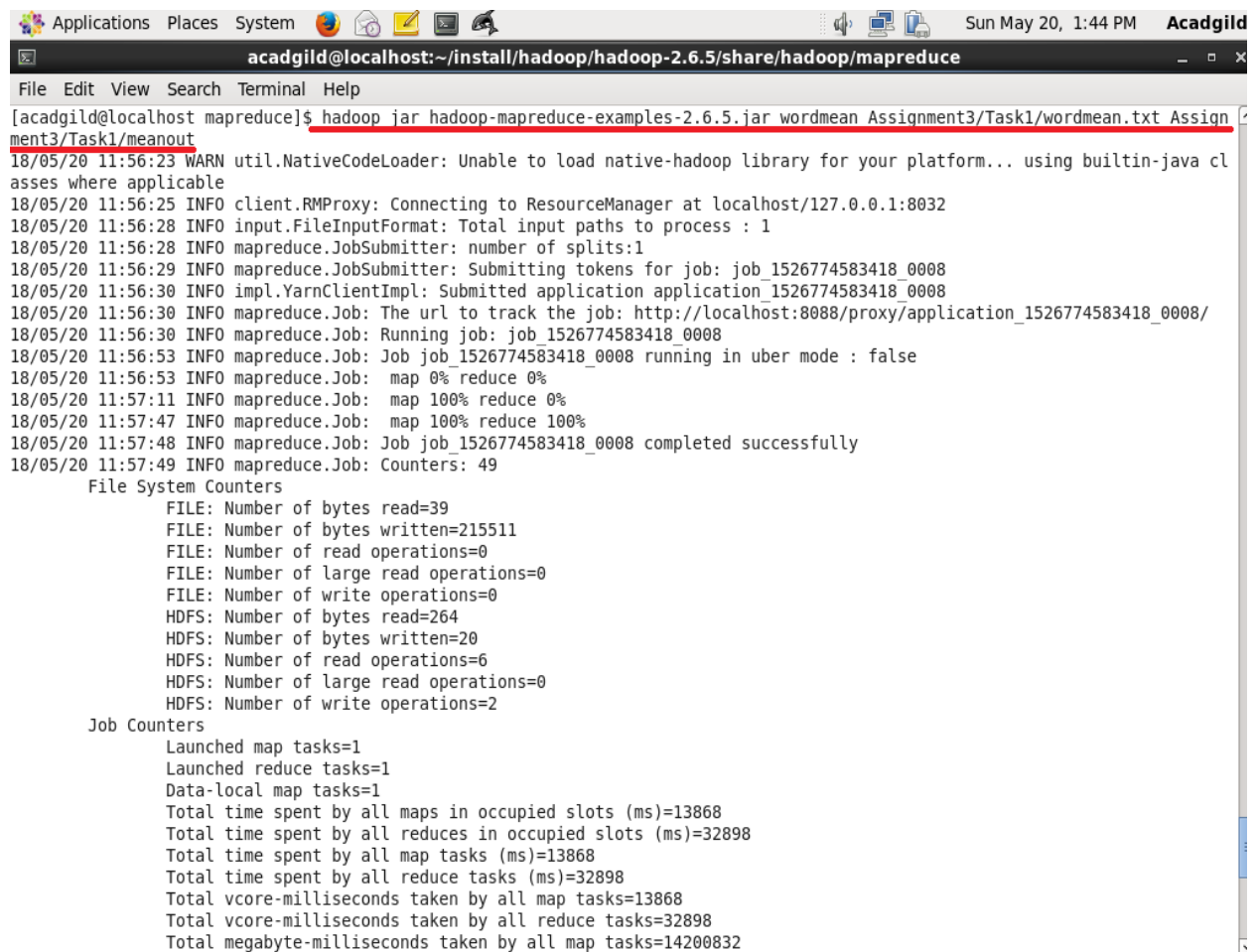Execute the jar file for word mean.

Syntax:

Hadoop jar <jarfilename> <mainclass name> <inputfile path> <outputFilePath>

Command:

Hadoop jar Hadoop-mapreduce-examples-2.5.6.jar wordmean Assignment3/Task1/wordmean.txt Assignment3/Task1/meanout

🖾   acadgild@localhost:~/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce   — □ ✕

File   Edit   View   Search   Terminal   Help

```
                Total vcore-milliseconds taken by all map tasks=13868
                Total vcore-milliseconds taken by all reduce tasks=32898
                Total megabyte-milliseconds taken by all map tasks=14200832
                Total megabyte-milliseconds taken by all reduce tasks=33687552
        Map-Reduce Framework
                Map input records=31
                Map output records=62
                Map output bytes=899
                Map output materialized bytes=39
                Input split bytes=131
                Combine input records=62
                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)=402
                CPU time spent (ms)=4990
                Physical memory (bytes) snapshot=292900864
                Virtual memory (bytes) snapshot=4121735168
                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=133
        File Output Format Counters
                Bytes Written=20
The mean is: 3.2903225806451615
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost mapreduce]$ █
```

🐧 [What is Hadoop? - De...   🖾 acadgild@localhost:~/   ☐ [task1_1.png]   🖾

Step3:

Display the contents of output file

🖾   acadgild@localhost:~/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce   — □

File   Edit   View   Search   Terminal   Help

```
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost mapreduce]$ hadoop fs -ls Assignment3/Task1/meanout
18/05/20 13:45:30 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
Found 2 items
-rw-r--r--   1 acadgild supergroup          0 2018-05-20 11:57 Assignment3/Task1/meanout/_SUCCESS
-rw-r--r--   1 acadgild supergroup         20 2018-05-20 11:57 Assignment3/Task1/meanout/part-r-00000
[acadgild@localhost mapreduce]$ hadoop fs -cat Assignment3/Task1/meanout/part-r-00000
18/05/20 13:45:55 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
count   31
length  102
[acadgild@localhost mapreduce]$ █
```
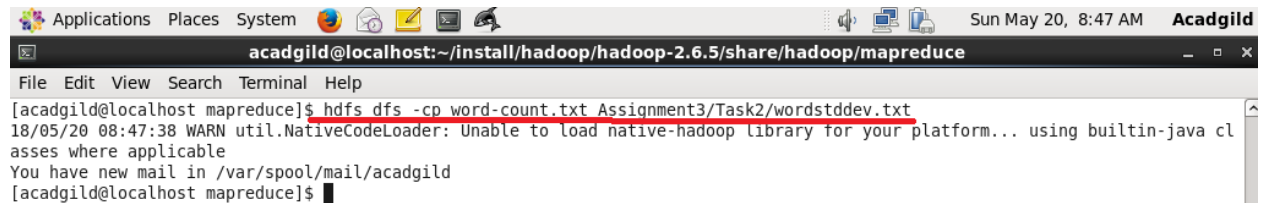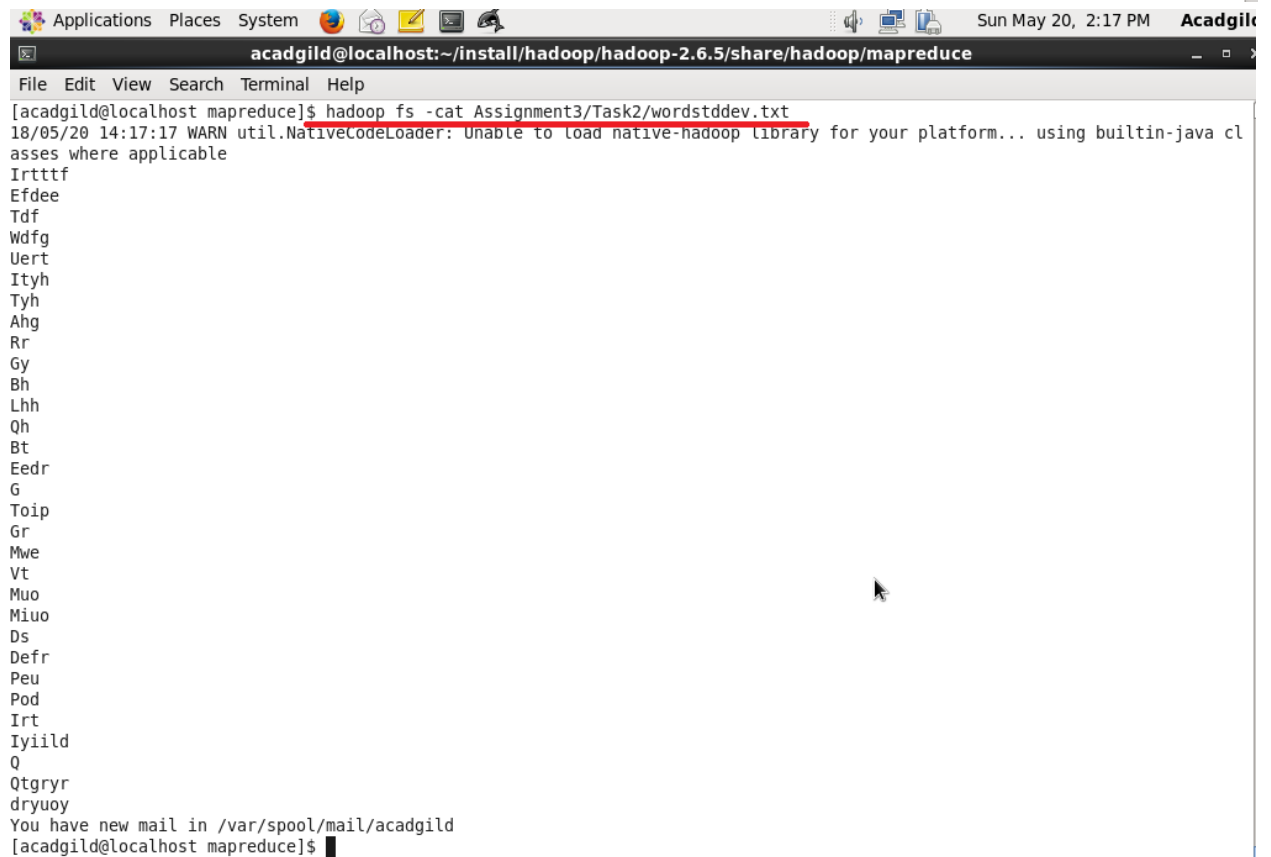
3. Word Standard Deviation

<u>Step1:</u>

Copy the file wordcount.txt to wordstddev.txt

Command:

Hdfs dfs -cp wordcount.txt Assignment3/Task2/wordstddev.txt

Step2:

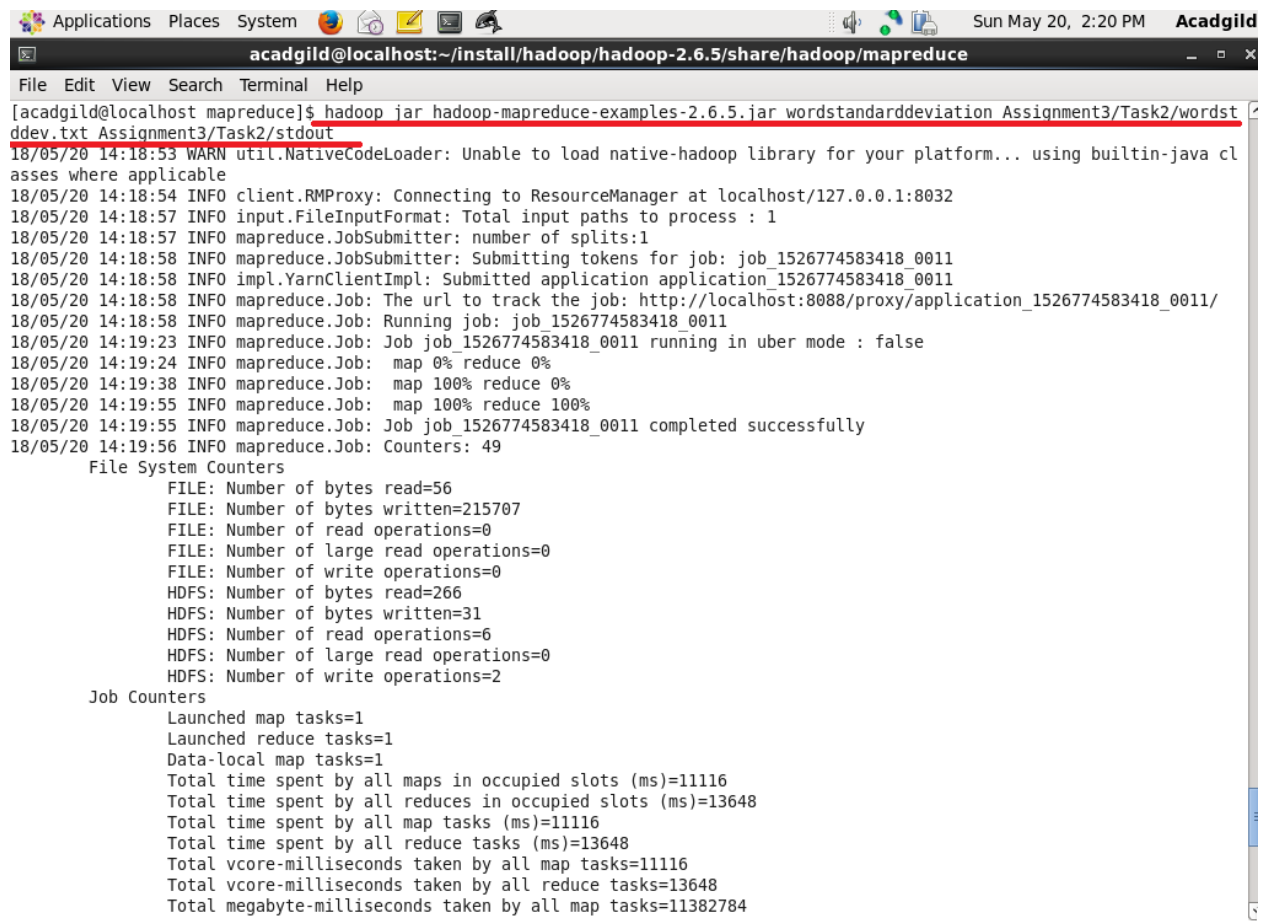Execute the jar file for word standard deviation.

Syntax:

Hadoop jar <jarfilename> <mainclass name> <inputfile path> <outputFilePath>

Command:

Hadoop jar Hadoop-mapreduce-examples-2.5.6.jar wordstandarddeviation Assignment3/Task2/wordstddev.txt Assignment3/Task2/stdout

File Edit View Search Terminal Help

```
                Total vcore-milliseconds taken by all map tasks=11116
                Total vcore-milliseconds taken by all reduce tasks=13648
                Total megabyte-milliseconds taken by all map tasks=11382784
                Total megabyte-milliseconds taken by all reduce tasks=13975552
        Map-Reduce Framework
                Map input records=31
                Map output records=93
                Map output bytes=1364
                Map output materialized bytes=56
                Input split bytes=133
                Combine input records=93
                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)=211
                CPU time spent (ms)=2430
                Physical memory (bytes) snapshot=293580800
                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=133
        File Output Format Counters
                Bytes Written=31
The standard deviation is: 1.3956972679215436
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost mapreduce]$ ▮
```

Step3:

Display the contents of output file

File Edit View Search Terminal Help

```
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost mapreduce]$ hadoop fs -ls Assignment3/Task2/stdout
18/05/20 14:21:22 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
Found 2 items
-rw-r--r--   1 acadgild supergroup          0 2018-05-20 14:19 Assignment3/Task2/stdout/_SUCCESS
-rw-r--r--   1 acadgild supergroup         31 2018-05-20 14:19 Assignment3/Task2/stdout/part-r-00000
[acadgild@localhost mapreduce]$ hadoop fs -cat Assignment3/Task2/stdout/part-r-00000
18/05/20 14:21:38 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
count   31
length  102
square  396
[acadgild@localhost mapreduce]$ ▮
```