

### **Assignment 3:**

Submitted by Jyoti Jha

DOS: 24<sup>th</sup> July'18

#### **Task 1:**

WordMedian WordMean WordStandardDeviation programs using hadoop-mapreduce-examples-2.9.0.jar file from path “/home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce”.

#### **Solution:**

#### **Input File Used:**

#### **Test.txt**

```
[acadgild@localhost ~]$ hadoop fs -cat /test.txt
18/07/24 21:45:08 WARN util.NativeCodeLoader: Unab
pplicable
Hello
Hello
Hi hello namastey
This is wonderful session.You have new mail in /va
[acadgild@localhost ~]$
```

NOTE: there was no jar file with 2.9.0 version; hence taken 2.6.5 instead.

#### **WordMedian:**

Counts the median length of the words in the input files.

```
[acadgild@localhost ~]$ hadoop jar /home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce/hadoop-mapreduce-examples-2.6.5.jar wordmedian /test.txt /output1
18/07/24 21:09:03 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/07/24 21:09:04 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/07/24 21:09:08 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
18/07/24 21:09:09 INFO input.FileInputFormat: Total input paths to process : 1
18/07/24 21:09:09 INFO mapreduce.JobSubmitter: number of splits:1
18/07/24 21:09:09 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1532444686738_0003
18/07/24 21:09:10 INFO impl.YarnClientImpl: Submitted application application_1532444686738_0003
18/07/24 21:09:10 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1532444686738_0003/
18/07/24 21:09:10 INFO mapreduce.Job: Running job: job_1532444686738_0003
18/07/24 21:09:37 INFO mapreduce.Job: Job job_1532444686738_0003 running in uber mode : false
18/07/24 21:09:37 INFO mapreduce.Job: map 0% reduce 0%
18/07/24 21:09:52 INFO mapreduce.Job: map 100% reduce 0%
18/07/24 21:10:08 INFO mapreduce.Job: map 100% reduce 100%
18/07/24 21:10:09 INFO mapreduce.Job: Job job_1532444686738_0003 completed successfully
18/07/24 21:10:12 INFO mapreduce.Job: Counters: 49
  File System Counters
    FILE: Number of bytes read=56
    FILE: Number of bytes written=215151
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=154
    HDFS: Number of bytes written=20
    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)=12218
    Total time spent by all reduces in occupied slots (ms)=13447
    Total time spent by all map tasks (ms)=12218
```

```

Total time spent by all reduce tasks (ms)=13447
Total vcore-milliseconds taken by all map tasks=12218
Total vcore-milliseconds taken by all reduce tasks=13447
Total megabyte-milliseconds taken by all map tasks=12511232
Total megabyte-milliseconds taken by all reduce tasks=13769728
Map-Reduce Framework
  Map input records=4
  Map output records=9
  Map output bytes=72
  Map output materialized bytes=56
  Input split bytes=95
  Combine input records=9
  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)=511
  CPU time spent (ms)=4030
  Physical memory (bytes) snapshot=297725952
  Virtual memory (bytes) snapshot=4118224896
  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=59
File Output Format Counters
  Bytes Written=20
The median is: 5
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$
```

## Output1:

```

[acadgild@localhost ~]$ hadoop fs -ls /Output1
18/07/24 21:14:53 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 2 items
-rw-r--r-- 1 acadgild supergroup          0 2018-07-24 21:10 /Output1/_SUCCESS
-rw-r--r-- 1 acadgild supergroup        20 2018-07-24 21:10 /Output1/part-r-000000
[acadgild@localhost ~]$ hadoop fs -cat /Output1/part-r-000000
18/07/24 21:15:10 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
2
2
4
1
5
3
8
2
9
1
r
```

## WordMean:

Counts the average length of the words in the input files.

```
[acadgild@localhost ~]$ hadoop jar /home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce/hadoop-mapreduce-examples-2.6.5.jar wordmean /test.txt /Output2
18/07/24 21:17:24 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/07/24 21:17:26 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/07/24 21:17:31 INFO input.FileInputFormat: Total input paths to process : 1
18/07/24 21:17:31 INFO mapreduce.JobSubmitter: number of splits:1
18/07/24 21:17:31 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1532444686738_0004
18/07/24 21:17:32 INFO impl.YarnClientImpl: Submitted application application_1532444686738_0004
18/07/24 21:17:32 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1532444686738_0004/
18/07/24 21:17:32 INFO mapreduce.Job: Running job: job_1532444686738_0004
18/07/24 21:17:54 INFO mapreduce.Job: Job job_1532444686738_0004 running in uber mode : false
18/07/24 21:17:54 INFO mapreduce.Job: map 0% reduce 0%
18/07/24 21:18:09 INFO mapreduce.Job: map 100% reduce 0%
18/07/24 21:18:24 INFO mapreduce.Job: map 100% reduce 100%
18/07/24 21:18:26 INFO mapreduce.Job: Job job_1532444686738_0004 completed successfully
18/07/24 21:18:26 INFO mapreduce.Job: Counters: 49
  File System Counters
    FILE: Number of bytes read=39
    FILE: Number of bytes written=215381
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=154
    HDFS: Number of bytes written=18
    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)=11534
    Total time spent by all reduces in occupied slots (ms)=12213
    Total time spent by all map tasks (ms)=11534
    Total time spent by all reduce tasks (ms)=12213
    Total vcore-milliseconds taken by all map tasks=11534
    Total vcore-milliseconds taken by all reduce tasks=12213
    Total megabyte-milliseconds taken by all map tasks=11810816
```

```

Total time spent by all reduce tasks (ms)=12213
Total vcore-milliseconds taken by all map tasks=11534
Total vcore-milliseconds taken by all reduce tasks=12213
Total megabyte-milliseconds taken by all map tasks=11810816
Total megabyte-milliseconds taken by all reduce tasks=12506112
Map-Reduce Framework
  Map input records=4
  Map output records=18
  Map output bytes=261
  Map output materialized bytes=39
  Input split bytes=95
  Combine input records=18
  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)=280
  CPU time spent (ms)=3810
  Physical memory (bytes) snapshot=296247296
  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=59
File Output Format Counters
  Bytes Written=18
The mean is: 5.333333333333333
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$ █

```

## Output2:

```

[acadgild@localhost ~]$ hadoop fs -cat /Output2/part-r-00000
18/07/24 21:24:15 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
count      9
length    48
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$ █

```

## WordStandardDeviation:

Counts the standard deviation of the length of the words in the input files.

```
[acadgild@localhost ~]$ hadoop jar /home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce/hadoop-mapreduce-examples-2.6.5.jar wordstandarddeviation /test.txt /Output3
18/07/24 21:27:52 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/07/24 21:27:56 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/07/24 21:27:58 INFO input.FileInputFormat: Total input paths to process : 1
18/07/24 21:27:58 INFO mapreduce.JobSubmitter: number of splits:1
18/07/24 21:27:58 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1532444686738_0005
18/07/24 21:27:59 INFO impl.YarnClientImpl: Submitted application application_1532444686738_0005
18/07/24 21:27:59 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1532444686738_0005/
18/07/24 21:27:59 INFO mapreduce.Job: Running job: job_1532444686738_0005
18/07/24 21:28:19 INFO mapreduce.Job: Job job_1532444686738_0005 running in uber mode : false
18/07/24 21:28:19 INFO mapreduce.Job: map 0% reduce 0%
18/07/24 21:28:31 INFO mapreduce.Job: map 100% reduce 0%
18/07/24 21:28:44 INFO mapreduce.Job: map 100% reduce 100%
18/07/24 21:28:44 INFO mapreduce.Job: Job job_1532444686738_0005 completed successfully
18/07/24 21:28:44 INFO mapreduce.Job: Counters: 49
    File System Counters
      FILE: Number of bytes read=56
      FILE: Number of bytes written=215575
      FILE: Number of read operations=0
      FILE: Number of large read operations=0
      FILE: Number of write operations=0
      HDFS: Number of bytes read=154
      HDFS: Number of bytes written=29
      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)=8391
      Total time spent by all reduces in occupied slots (ms)=9781
      Total time spent by all map tasks (ms)=8391
      Total time spent by all reduce tasks (ms)=9781
      Total vcore-milliseconds taken by all map tasks=8391
      Total vcore-milliseconds taken by all reduce tasks=9781
      Total megabyte-milliseconds taken by all map tasks=8592384
      Total megabyte-milliseconds taken by all reduce tasks=10015744
```

```

Total time spent by all reduce tasks (ms)=9781
Total vcore-milliseconds taken by all map tasks=8391
Total vcore-milliseconds taken by all reduce tasks=9781
Total megabyte-milliseconds taken by all map tasks=8592384
Total megabyte-milliseconds taken by all reduce tasks=10015744
Map-Reduce Framework
  Map input records=4
  Map output records=27
  Map output bytes=396
  Map output materialized bytes=56
  Input split bytes=95
  Combine input records=27
  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)=325
  CPU time spent (ms)=1890
  Physical memory (bytes) snapshot=300302336
  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=59
File Output Format Counters
  Bytes Written=29
The standard deviation is: 2.4037008503093262
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$

```

### Output3:

```

[acadgild@localhost ~]$ hadoop fs -cat /Output3/part-r-00000
18/07/24 21:34:15 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
count 9
length 48
square 308
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$

```