

# ACADGILD

## Assignment 3.1

YARN – WordMedian, WordMean, WordStandardDeviation  
MR job execution.

**Ashish Kumar**  
**ashishashuu0602@gmail.com**



## Assignment 3.1

**TASK 1:** Execute WordMedian, WordMean and WordStandarddeviation using jar files present in VM.

**Solution:**

### WORDMEDIAN

Command – to execute MapReduce Jar –

```
hadoop jar /home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce/hadoop-mapreduce-examples-2.6.5.jar wordmedian /test.txt /wordmedout
```

where –

test.txt – input file used.

wordmedout – output file name.

```
[acadgild@192 ~]$ hadoop jar /home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce/hadoop-mapreduce-examples-2.6.5.jar wordmedian /test.txt /wordmedout
18/07/22 01:20:16 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/07/22 01:20:16 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/07/22 01:20:18 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
18/07/22 01:20:19 INFO input.FileInputFormat: Total input paths to process : 1
18/07/22 01:20:19 INFO mapreduce.JobSubmitter: number of splits:1
18/07/22 01:20:19 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1532202396031_0001
18/07/22 01:20:20 INFO impl.YarnClientImpl: Submitted application application_1532202396031_0001
18/07/22 01:20:20 INFO mapreduce.Job: The url to track the job: http://192.168.0.5:8088/proxy/application_1532202396031_0001/
18/07/22 01:20:20 INFO mapreduce.Job: Running job: job_1532202396031_0001
18/07/22 01:20:29 INFO mapreduce.Job: Job job_1532202396031_0001 running in uber mode : false
18/07/22 01:20:29 INFO mapreduce.Job: map 0% reduce 0%
18/07/22 01:20:35 INFO mapreduce.Job: map 100% reduce 0%
18/07/22 01:20:42 INFO mapreduce.Job: map 100% reduce 100%
18/07/22 01:20:43 INFO mapreduce.Job: Job job_1532202396031_0001 completed successfully
18/07/22 01:20:43 INFO mapreduce.Job: Counters: 49
  File System Counters
    FILE: Number of bytes read=36
    FILE: Number of bytes written=215117
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=280
    HDFS: Number of bytes written=13
    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)=3804
    Total time spent by all reduces in occupied slots (ms)=3706
    Total time spent by all map tasks (ms)=3804
    Total time spent by all reduce tasks (ms)=3706
    Total vcore-milliseonds taken by all map tasks=3804
    Total vcore-milliseonds taken by all reduce tasks=3706
    Total megabyte-milliseonds taken by all map tasks=3895296
    Total megabyte-milliseonds taken by all reduce tasks=3794944
  Map-Reduce Framework
    Map input records=5
```

## Assignment 3.1

```
Map output records=20
Map output bytes=160
Map output materialized bytes=36
Input split bytes=95
Combine input records=20
Combine output records=3
Reduce input groups=3
Reduce shuffle bytes=36
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)=170
CPU time spent (ms)=1700
Physical memory (bytes) snapshot=423006208
Virtual memory (bytes) snapshot=4175822848
Total committed heap usage (bytes)=309854208

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=105
File Output Format Counters
  Bytes Written=13

The median is: 4
```

**wordmedout** – output file specified at the time of executing jar file.

Snip is below –

```
[acadgild@192 ~]$ hadoop fs -ls /wordmedout
18/07/22 01:22:13 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 2018-07-22 01:20 /wordmedout/_SUCCESS
-rw-r--r-- 1 acadgild supergroup 13 2018-07-22 01:20 /wordmedout/part-r-00000
[acadgild@192 ~]$ hadoop fs -ls /wordmedout/part-r-00000
18/07/22 01:22:25 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
-rw-r--r-- 1 acadgild supergroup 13 2018-07-22 01:20 /wordmedout/part-r-00000
[acadgild@192 ~]$ hadoop fs -cat /wordmedout/part-r-00000
18/07/22 01:22:37 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
2      5
4      10
7      5
[acadgild@192 ~]$
```

**Test.txt** –

```
[acadgild@192 ~]$ cat test.txt
This is BDHS session
This is BDHS session
This is BDHS session
This is BDHS session
This is BDHS session
```

```
[acadgild@192 ~]$ hadoop fs -cat /test.txt
18/07/22 01:46:04 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
This is BDHS session
This is BDHS session
This is BDHS session
This is BDHS session
This is BDHS session
[acadgild@192 ~]$
```

## WORDMEAN

**hadoop jar /home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce/hadoop-mapreduce-examples-2.6.5.jar wordmean /test.txt /wordmeanout**

```
[acadgild@192 ~]$ hadoop jar /home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce/hadoop-mapreduce-examples-2.6.5.jar wordmean /test.txt /wordmeanout
18/07/22 02:00:24 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/07/22 02:00:24 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/07/22 02:00:25 INFO input.FileInputFormat: Total input paths to process : 1
18/07/22 02:00:26 INFO mapreduce.JobSubmitter: number of splits:1
18/07/22 02:00:26 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1532202396031_0002
18/07/22 02:00:28 INFO impl.YarnClientImpl: Submitted application application_1532202396031_0002
18/07/22 02:00:28 INFO mapreduce.Job: The url to track the job: http://192.168.0.5:8088/proxy/application_1532202396031_0002/
18/07/22 02:00:28 INFO mapreduce.Job: Running job: job_1532202396031_0002
18/07/22 02:00:36 INFO mapreduce.Job: Job job_1532202396031_0002 running in uber mode : false
18/07/22 02:00:36 INFO mapreduce.Job: map 0% reduce 0%
18/07/22 02:00:45 INFO mapreduce.Job: map 100% reduce 0%
18/07/22 02:00:52 INFO mapreduce.Job: map 100% reduce 100%
18/07/22 02:00:54 INFO mapreduce.Job: Job job_1532202396031_0002 completed successfully
18/07/22 02:00:54 INFO mapreduce.Job: Counters: 49

  File System Counters
    FILE: Number of bytes read=39
    FILE: Number of bytes written=215389
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=200
    HDFS: Number of bytes written=19
    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)=6789
    Total time spent by all reduces in occupied slots (ms)=4220
    Total time spent by all map tasks (ms)=6789
    Total time spent by all reduce tasks (ms)=4220
    Total vcore-milliseonds taken by all map tasks=6789
    Total vcore-milliseonds taken by all reduce tasks=4220
    Total megabyte-milliseonds taken by all map tasks=6951936
    Total megabyte-milliseonds taken by all reduce tasks=4321280

  Map-Reduce Framework
    Map input records=5
    Map output records=40
```

```
Combine input records=40
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)=168
CPU time spent (ms)=1660
Physical memory (bytes) snapshot=428376064
Virtual memory (bytes) snapshot=4167491584
Total committed heap usage (bytes)=306708480

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=105
File Output Format Counters
  Bytes Written=19

The mean is: 4.25
[acadgild@192 ~]$
```

The Mean is : 4.25

## Assignment 3.1

### Output –

```
[acadgild@192 ~]$ hadoop fs -ls /wordmeanout
18/07/22 02:05:11 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 2018-07-22 02:00 /wordmeanout/ SUCCESS
-rw-r--r-- 1 acadgild supergroup        19 2018-07-22 02:00 /wordmeanout/part-r-00000
[acadgild@192 ~]$ hadoop fs -ls /wordmeanout/part-r-00000
18/07/22 02:05:21 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
-rw-r--r-- 1 acadgild supergroup        19 2018-07-22 02:00 /wordmeanout/part-r-00000
[acadgild@192 ~]$ hadoop fs -cat /wordmeanout/part-r-00000
18/07/22 02:05:39 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
count 20
length 85
[acadgild@192 ~]$
```

## WordStandardDeviation

**hadoop jar /home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce/hadoop-mapreduce-examples-2.6.5.jar wordstandarddeviation /test.txt /wordstandrddevout**

```
[acadgild@192 ~]$ hadoop jar /home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce/hadoop-mapreduce-examples-2.6.5.jar wordstandarddeviation /test.txt /wordstandrddevout
18/07/22 02:13:21 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/07/22 02:13:22 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/07/22 02:13:26 INFO input.FileInputFormat: Total input paths to process : 1
18/07/22 02:13:26 INFO mapreduce.JobSubmitter: number of splits:1
18/07/22 02:13:27 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1532202396031_0003
18/07/22 02:13:27 INFO impl.YarnClientImpl: Submitted application application_1532202396031_0003
18/07/22 02:13:27 INFO mapreduce.Job: The url to track the job: http://192.168.0.5:8088/proxy/application_1532202396031_0003/
18/07/22 02:13:27 INFO mapreduce.Job: Running job: job_1532202396031_0003
18/07/22 02:13:35 INFO mapreduce.Job: Job job_1532202396031_0003 running in uber mode : false
18/07/22 02:13:35 INFO mapreduce.Job: map 0% reduce 0%
18/07/22 02:13:42 INFO mapreduce.Job: map 100% reduce 0%
18/07/22 02:13:53 INFO mapreduce.Job: map 100% reduce 100%
18/07/22 02:13:58 INFO mapreduce.Job: Job job_1532202396031_0003 completed successfully
18/07/22 02:13:58 INFO mapreduce.Job: Counters: 49
  File System Counters
    FILE: Number of bytes read=56
    FILE: Number of bytes written=215595
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=200
    HDFS: Number of bytes written=30
    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)=5150
    Total time spent by all reduces in occupied slots (ms)=8303
    Total time spent by all map tasks (ms)=5150
    Total time spent by all reduce tasks (ms)=8303
    Total vcore-milliseconds taken by all map tasks=5150
    Total vcore-milliseconds taken by all reduce tasks=8303
    Total megabyte-milliseconds taken by all map tasks=5273600
    Total megabyte-milliseconds taken by all reduce tasks=8502272
  Map-Reduce Framework
```

## Assignment 3.1

```
Map-Reduce Framework
  Map input records=5
  Map output records=60
  Map output bytes=880
  Map output materialized bytes=56
  Input split bytes=95
  Combine input records=60
  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)=120
  CPU time spent (ms)=1530
  Physical memory (bytes) snapshot=425201664
  Virtual memory (bytes) snapshot=4175671296
  Total committed heap usage (bytes)=310378496

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=105
File Output Format Counters
  Bytes Written=30
The standard deviation is: 1.7853571071357126
```

The Standard Deviation is: **1.7853571071357126**

### Output –

```
[acadgild@192 ~]$ hadoop fs -ls /wordstandrddevout
18/07/22 02:18:48 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform
Found 2 items
-rw-r--r--  1 acadgild supergroup          0 2018-07-22 02:13 /wordstandrddevout/_SUCCESS
-rw-r--r--  1 acadgild supergroup        30 2018-07-22 02:13 /wordstandrddevout/part-r-00000
[acadgild@192 ~]$ hadoop fs -cat /wordstandrddevout/part-r-00000
18/07/22 02:19:06 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform
count  20
length  85
square 425
[acadgild@192 ~]$
```

Input File used across execution of 3 Jars is as below –

test.txt –

```
[acadgild@192 ~]$ cat test.txt
This is BDHS session
This is BDHS session
This is BDHS session
This is BDHS session
This is BDHS session
```

```
[acadgild@192 ~]$ hadoop fs -cat /test.txt
18/07/22 01:46:04 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
This is BDHS session
This is BDHS session
This is BDHS session
This is BDHS session
This is BDHS session
[acadgild@192 ~]$
```