

Task 1:

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

Execution:

```
[acadgild@localhost ~]$ jps
3356 Jps
[acadgild@localhost ~]$ sudo service sshd start
[sudo] password for acadgild:
[acadgild@localhost ~]$ start-all.sh
This script is Deprecated. Instead use start-dfs.sh and start-yarn.sh
18/06/25 00:30:40 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your
platform... using builtin-java classes where applicable
Starting namenodes on [localhost]
localhost: starting namenode, logging to
/home/acadgild/install/hadoop/hadoop-2.6.5/logs/hadoop-acadgild-namenode-localhost.localdomain.ou
t
localhost: starting datanode, logging to
/home/acadgild/install/hadoop/hadoop-2.6.5/logs/hadoop-acadgild-datanode-localhost.localdomain.out
Starting secondary namenodes [0.0.0.0]
0.0.0.0: starting secondarynamenode, logging to
/home/acadgild/install/hadoop/hadoop-2.6.5/logs/hadoop-acadgild-secondarynamenode-localhost.local
domain.out
18/06/25 00:31:24 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your
platform... using builtin-java classes where applicable
starting yarn daemons
starting resourcemanager, logging to
/home/acadgild/install/hadoop/hadoop-2.6.5/logs/yarn-acadgild-resourcemanager-localhost.localdomai
n.out
localhost: starting nodemanager, logging to
/home/acadgild/install/hadoop/hadoop-2.6.5/logs/yarn-acadgild-nodemanager-localhost.localdomain.ou
t
[acadgild@localhost ~]$ jps
3600 DataNode
4244 Jps
4036 NodeManager
3759 SecondaryNameNode
3503 NameNode
3935 ResourceManager
[acadgild@localhost ~]$ pwd
/home/acadgild
[acadgild@localhost ~]$ ls
Desktop  eclipse      install    Music    Templates
Documents eclipse-workspace IntelliJ  Pictures tweets
Downloads IdeaProjects metastore_db Public   Videos
[acadgild@localhost ~]$ cd Desktop
```

```
[acadgild@localhost Desktop]$ nano file.txt
```

```
[acadgild@localhost Desktop]$ cat file.txt
```

Hadoop is an open-source software framework for storing data and running applications on clusters of commodity hardware.

It provides massive storage for any kind of data, enormous processing power and the ability to handle virtually limitless concurrent tasks or jobs.

One such project was an open-source web search engine called Nutch – the brainchild of Doug Cutting and Mike Cafarella. They wanted to return web search results faster by distributing data and calculations across different computers so multiple tasks could be accomplished simultaneously.

During this time, another search engine project called Google was in progress. It was based on the same concept – storing and processing data in a distributed, automated way so that relevant web search results could be returned faster.

In 2006, Cutting joined Yahoo and took with him the Nutch project as well as ideas based on Google's early work with automating distributed data storage and processing.

The Nutch project was divided – the web crawler portion remained as Nutch and the distributed computing and processing portion became Hadoop (named after Cutting's son's toy elephant).

In 2008, Yahoo released Hadoop as an open-source project.

Today, Hadoop's framework and ecosystem of technologies are managed and maintained by the non-profit Apache Software Foundation (ASF), a global community of software developers and contributors.

```
[acadgild@localhost Desktop]$ pwd
```

```
/home/acadgild/Desktop
```

```
[acadgild@localhost Desktop]$ hdfs dfs -put /home/acadgild/Desktop/file.txt
```

```
/user/acadgild/hadoopdata/assignment3
```

```
18/06/25 01:06:37 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
```

```
[acadgild@localhost Desktop]$ hdfs dfs -ls /user/acadgild/hadoopdata/assignment3
```

```
18/06/25 01:07:09 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
```

```
Found 1 items
```

```
-rw-r--r--  1 acadgild supergroup    1417 2018-06-25 01:06
```

```
/user/acadgild/hadoopdata/assignment3/file.txt
```

1. Word Median

```
[acadgild@localhost Desktop]$ hadoop jar
```

```
/home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce/hadoop-mapreduce-examples-2.6.5.jar wordmedian /user/acadgild/hadoopdata/assignment3/file.txt
```

```
/user/acadgild/hadoopdata/assignment3/WordMedianOutput
```

```
18/06/25 01:26:04 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
```

```
18/06/25 01:26:06 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
```

18/06/25 01:26:07 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
18/06/25 01:26:08 INFO input.FileInputFormat: Total input paths to process : 1
18/06/25 01:26:08 INFO mapreduce.JobSubmitter: number of splits:1
18/06/25 01:26:09 INFO mapreduce.JobSubmitter: Submitting tokens for job:
job_1529866896407_0002
18/06/25 01:26:09 INFO impl.YarnClientImpl: Submitted application
application_1529866896407_0002
18/06/25 01:26:10 INFO mapreduce.Job: The url to track the job:
http://localhost:8088/proxy/application_1529866896407_0002/
18/06/25 01:26:10 INFO mapreduce.Job: Running job: job_1529866896407_0002
18/06/25 01:26:27 INFO mapreduce.Job: Job job_1529866896407_0002 running in uber mode : false
18/06/25 01:26:27 INFO mapreduce.Job: map 0% reduce 0%
18/06/25 01:26:37 INFO mapreduce.Job: map 100% reduce 0%
18/06/25 01:26:48 INFO mapreduce.Job: map 100% reduce 100%
18/06/25 01:26:49 INFO mapreduce.Job: Job job_1529866896407_0002 completed successfully
18/06/25 01:26:49 INFO mapreduce.Job: Counters: 49

File System Counters

FILE: Number of bytes read=146
FILE: Number of bytes written=215497
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=1549
HDFS: Number of bytes written=70
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)=7909
Total time spent by all reduces in occupied slots (ms)=7883
Total time spent by all map tasks (ms)=7909
Total time spent by all reduce tasks (ms)=7883
Total vcore-milliseconds taken by all map tasks=7909
Total vcore-milliseconds taken by all reduce tasks=7883
Total megabyte-milliseconds taken by all map tasks=8098816
Total megabyte-milliseconds taken by all reduce tasks=8072192

Map-Reduce Framework

Map input records=9
Map output records=213
Map output bytes=1704
Map output materialized bytes=146
Input split bytes=132
Combine input records=213
Combine output records=14
Reduce input groups=14

```

Reduce shuffle bytes=146
Reduce input records=14
Reduce output records=14
Spilled Records=28
Shuffled Maps =1
Failed Shuffles=0
Merged Map outputs=1
GC time elapsed (ms)=308
CPU time spent (ms)=2110
Physical memory (bytes) snapshot=236564480
Virtual memory (bytes) snapshot=681779200
Total committed heap usage (bytes)=140042240

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=1417

File Output Format Counters
  Bytes Written=70

```

The median is: 5

```
[acadgild@localhost Desktop]$ hdfs dfs -ls
```

```
/user/acadgild/hadoopdata/assignment3/WordMedianOutput18/06/25 01:27:05 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 2018-06-25 01:26
```

```
/user/acadgild/hadoopdata/assignment3/WordMedianOutput/_SUCCESS
```

```
-rw-r--r--  1 acadgild supergroup    70 2018-06-25 01:26
```

```
/user/acadgild/hadoopdata/assignment3/WordMedianOutput/part-r-00000
```

```
[acadgild@localhost Desktop]$ hdfs dfs -cat
```

```
/user/acadgild/hadoopdata/assignment3/WordMedianOutput/part-r-0000018/06/25 01:27:39 WARN
```

```
util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java
classes where applicable
```

```

1      5
2     31
3     36
4     17
5     22
6     25
7     23
8     14
9     14
10    12
11     6
12     6

```

13 1
15 1

2. Word Mean:

```
[acadgild@localhost Desktop]$ hadoop jar
/home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce/hadoop-mapreduce-examples-2.
6.5.jar wordmean /user/acadgild/hadoopdata/assignment3/file.txt
/user/acadgild/hadoopdata/assignment3/WrdMeanOutput
18/06/25 01:28:54 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your
platform... using builtin-java classes where applicable
18/06/25 01:28:56 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/06/25 01:28:58 INFO input.FileInputFormat: Total input paths to process : 1
18/06/25 01:28:58 INFO mapreduce.JobSubmitter: number of splits:1
18/06/25 01:28:58 INFO mapreduce.JobSubmitter: Submitting tokens for job:
job_1529866896407_0003
18/06/25 01:28:59 INFO impl.YarnClientImpl: Submitted application
application_1529866896407_0003
18/06/25 01:28:59 INFO mapreduce.Job: The url to track the job:
http://localhost:8088/proxy/application_1529866896407_0003/
18/06/25 01:28:59 INFO mapreduce.Job: Running job: job_1529866896407_0003
18/06/25 01:29:12 INFO mapreduce.Job: Job job_1529866896407_0003 running in uber mode : false
18/06/25 01:29:12 INFO mapreduce.Job: map 0% reduce 0%
18/06/25 01:29:21 INFO mapreduce.Job: map 100% reduce 0%
18/06/25 01:29:31 INFO mapreduce.Job: map 100% reduce 100%
18/06/25 01:29:31 INFO mapreduce.Job: Job job_1529866896407_0003 completed successfully
18/06/25 01:29:31 INFO mapreduce.Job: Counters: 49
    File System Counters
        FILE: Number of bytes read=39
        FILE: Number of bytes written=215541
        FILE: Number of read operations=0
        FILE: Number of large read operations=0
        FILE: Number of write operations=0
        HDFS: Number of bytes read=1549
        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)=6939
        Total time spent by all reduces in occupied slots (ms)=7289
        Total time spent by all map tasks (ms)=6939
        Total time spent by all reduce tasks (ms)=7289
        Total vcore-milliseconds taken by all map tasks=6939
        Total vcore-milliseconds taken by all reduce tasks=7289
```

Total megabyte-milliseconds taken by all map tasks=7105536
Total megabyte-milliseconds taken by all reduce tasks=7463936

Map-Reduce Framework

Map input records=9
Map output records=426
Map output bytes=6177
Map output materialized bytes=39
Input split bytes=132
Combine input records=426
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)=283
CPU time spent (ms)=2100
Physical memory (bytes) snapshot=236646400
Virtual memory (bytes) snapshot=681779200
Total committed heap usage (bytes)=139964416

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=1417

File Output Format Counters

Bytes Written=22

The mean is: 5.577464788732394

```
[acadgild@localhost Desktop]$ hdfs dfs -ls
```

```
/user/acadgild/hadoopdata/assignment3/WrdMeanOutput18/06/25 01:30:23 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 2018-06-25 01:29
```

```
/user/acadgild/hadoopdata/assignment3/WrdMeanOutput/_SUCCESS
```

```
-rw-r--r--  1 acadgild supergroup    22 2018-06-25 01:29
```

```
/user/acadgild/hadoopdata/assignment3/WrdMeanOutput/part-r-00000
```

```
[acadgild@localhost Desktop]$ hdfs dfs -cat
```

```
/user/acadgild/hadoopdata/assignment3/WrdMeanOutput/part-r-0000018/06/25 01:31:06 WARN
```

```
util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java  
classes where applicable
```

```
count  213
```

length 1188

3. WordStandardDeviation :

```
[acadgild@localhost Desktop]$ hadoop jar
/home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce/hadoop-mapreduce-examples-2.
6.5.jar wordstandarddeviation /user/acadgild/hadoopdata/assignment3/file.txt
/user/acadgild/hadoopdata/assignment3/WrdSDoutput
18/06/25 01:33:27 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your
platform... using builtin-java classes where applicable
18/06/25 01:33:29 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/06/25 01:33:31 INFO input.FileInputFormat: Total input paths to process : 1
18/06/25 01:33:31 INFO mapreduce.JobSubmitter: number of splits:1
18/06/25 01:33:31 INFO mapreduce.JobSubmitter: Submitting tokens for job:
job_1529866896407_0004
18/06/25 01:33:32 INFO impl.YarnClientImpl: Submitted application
application_1529866896407_0004
18/06/25 01:33:32 INFO mapreduce.Job: The url to track the job:
http://localhost:8088/proxy/application_1529866896407_0004/
18/06/25 01:33:32 INFO mapreduce.Job: Running job: job_1529866896407_0004
18/06/25 01:33:44 INFO mapreduce.Job: Job job_1529866896407_0004 running in uber mode : false
18/06/25 01:33:44 INFO mapreduce.Job: map 0% reduce 0%
18/06/25 01:33:54 INFO mapreduce.Job: map 100% reduce 0%
18/06/25 01:34:04 INFO mapreduce.Job: map 100% reduce 100%
18/06/25 01:34:05 INFO mapreduce.Job: Job job_1529866896407_0004 completed successfully
18/06/25 01:34:05 INFO mapreduce.Job: Counters: 49
    File System Counters
        FILE: Number of bytes read=56
        FILE: Number of bytes written=215731
        FILE: Number of read operations=0
        FILE: Number of large read operations=0
        FILE: Number of write operations=0
        HDFS: Number of bytes read=1549
        HDFS: Number of bytes written=34
        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)=6603
        Total time spent by all reduces in occupied slots (ms)=7759
        Total time spent by all map tasks (ms)=6603
        Total time spent by all reduce tasks (ms)=7759
        Total vcore-milliseconds taken by all map tasks=6603
        Total vcore-milliseconds taken by all reduce tasks=7759
        Total megabyte-milliseconds taken by all map tasks=6761472
```

Total megabyte-milliseconds taken by all reduce tasks=7945216

Map-Reduce Framework

- Map input records=9
- Map output records=639
- Map output bytes=9372
- Map output materialized bytes=56
- Input split bytes=132
- Combine input records=639
- 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)=288
- CPU time spent (ms)=2170
- Physical memory (bytes) snapshot=236331008
- Virtual memory (bytes) snapshot=681779200
- Total committed heap usage (bytes)=139988992

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=1417

File Output Format Counters

- Bytes Written=34

The standard deviation is: 2.9787761224582323

```
[acadgild@localhost Desktop]$ hdfs dfs -ls /user/acadgild/hadoopdata/assignment3/WrdSDoutput
18/06/25 01:36:37 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 2018-06-25 01:34
/user/acadgild/hadoopdata/assignment3/WrdSDoutput/_SUCCESS
-rw-r--r--  1 acadgild supergroup    34 2018-06-25 01:34
/user/acadgild/hadoopdata/assignment3/WrdSDoutput/part-r-00000
```

```
[acadgild@localhost Desktop]$ hdfs dfs -cat
/user/acadgild/hadoopdata/assignment3/WrdSDoutput/part-r-00000
18/06/25 01:38:48 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your
platform... using builtin-java classes where applicable
```


count 213
length 1188
square 8516