

## Vijay's Assignment 3

### Task 1:

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

Refer path below.

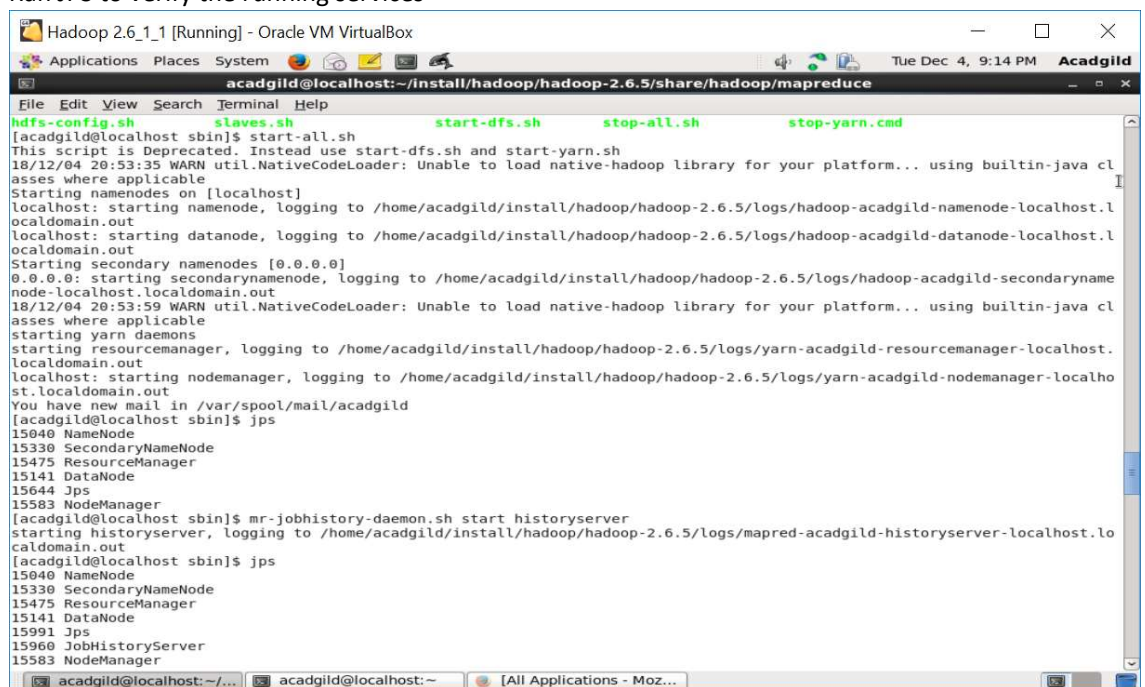
/home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce.

**I have added wordcount output also in my assignment and provided the explanation for it.**

**Rest of the commands I have given the associated screenshots alone**

Description of what was done

- 1) Started all the Hadoop services using start-all.sh
- 2) Started the JobHistoryServer using mr-jobhistory-daemon.sh
- 3) Ran JPS to verify the running services



```
Hadoop 2.6.1_1 [Running] - Oracle VM VirtualBox
Applications Places System
acadmild@localhost:~/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce
File Edit View Search Terminal Help
hdfs-config.sh slaves.sh start-dfs.sh stop-all.sh stop-yarn.cmd
[acadmild@localhost sbin]$ start-all.sh
This script is Deprecated. Instead use start-dfs.sh and start-yarn.sh
18/12/04 20:53:35 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
Starting namenodes on [localhost]
localhost: starting namenode, logging to /home/acadmild/install/hadoop/hadoop-2.6.5/logs/hadoop-acadmild-namenode-localhost.l
ocaldomain.out
localhost: starting datanode, logging to /home/acadmild/install/hadoop/hadoop-2.6.5/logs/hadoop-acadmild-datanode-localhost.l
ocaldomain.out
Starting secondary namenodes [0.0.0.0]
0.0.0.0: starting secondarynamenode, logging to /home/acadmild/install/hadoop/hadoop-2.6.5/logs/hadoop-acadmild-secondaryname
node-localhost.localdomain.out
18/12/04 20:53:59 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
starting yarn daemons
starting resourcemanager, logging to /home/acadmild/install/hadoop/hadoop-2.6.5/logs/yarn-acadmild-resourcemanager-localhost.
localdomain.out
localhost: starting nodemanager, logging to /home/acadmild/install/hadoop/hadoop-2.6.5/logs/yarn-acadmild-nodemanager-localhost.localdomain.out
You have new mail in /var/spool/mail/acadmild
[acadmild@localhost sbin]$ jps
15040 NameNode
15330 SecondaryNameNode
15475 ResourceManager
15141 DataNode
15644 Jps
15583 NodeManager
[acadmild@localhost sbin]$ mr-jobhistory-daemon.sh start historyserver
starting historyserver, logging to /home/acadmild/install/hadoop/hadoop-2.6.5/logs/mapred-acadmild-historyserver-localhost.localdomain.out
[acadmild@localhost sbin]$ jps
15040 NameNode
15330 SecondaryNameNode
15475 ResourceManager
15141 DataNode
15991 Jps
15960 JobHistoryServer
15583 NodeManager
```

- 4) Opened the Yarn webpage using localhost:8088/cluster link in Firefox
- 5) Changed my current working directory to /home/acadmild/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce
- 6) Ran the wordcount mapreduce job using Hadoop jar command. The input file was the one created in previous session at the location /user/acadmild/hadoop/word-count.txt file and the output directory was specified as /user/acadmild/hadoop/word-count-out
- 7) This triggered the job id 1543937045769\_001 and the same was monitored using Yarn webpage

```
Hadoop 2.6_1_1 [Running] - Oracle VM VirtualBox
Applications Places System
acagild@localhost:~/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce
File Edit View Search Terminal Help
[acagild@localhost mapreduce]$ hadoop jar hadoop-mapreduce-examples-2.6.5.jar wordcount /user/acagild/hadoop/word-count.txt
18/12/04 21:08:19 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
18/12/04 21:08:20 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/12/04 21:08:23 INFO input.FileInputFormat: Total input paths to process : 1
18/12/04 21:08:23 INFO mapreduce.JobSubmitter: number of splits:1
18/12/04 21:08:23 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1543937045769_0001
18/12/04 21:08:24 INFO impl.YarnClientImpl: Submitted application application_1543937045769_0001
18/12/04 21:08:24 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1543937045769_0001/
18/12/04 21:08:24 INFO mapreduce.Job: Running job: job_1543937045769_0001
18/12/04 21:08:40 INFO mapreduce.Job: Job job_1543937045769_0001 running in uber mode : false
18/12/04 21:08:40 INFO mapreduce.Job: map 0% reduce 0%
18/12/04 21:08:49 INFO mapreduce.Job: map 100% reduce 0%
18/12/04 21:08:58 INFO mapreduce.Job: map 100% reduce 100%
18/12/04 21:08:59 INFO mapreduce.Job: Job job_1543937045769_0001 completed successfully
18/12/04 21:08:59 INFO mapreduce.Job: Counters: 49
File System Counters
  FILE: Number of bytes read=442
  FILE: Number of bytes written=216297
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=369
  HDFS: Number of bytes written=284
  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)=5638
  Total time spent by all reduces in occupied slots (ms)=6389
  Total time spent by all map tasks (ms)=5638
  Total time spent by all reduce tasks (ms)=6389
  Total vcore-milliseconds taken by all map tasks=5638
  Total vcore-milliseconds taken by all reduce tasks=6389
  Total megabyte-milliseconds taken by all map tasks=5773312
```

Hadoop 2.6\_1\_1 [Running] - Oracle VM VirtualBox

All Applications - Mozilla Firefox

localhost:8088/cluster

Most Visited Getting Started yarn

### hadoop

## All Applications

Cluster

- About
- Nodes
- Applications
  - NEW
  - NEW SAVING
  - SUBMITTED
  - ACCEPTED
  - RUNNING
  - FINISHED
  - FAILED
  - KILLED
- Scheduler

Tools

Cluster Metrics

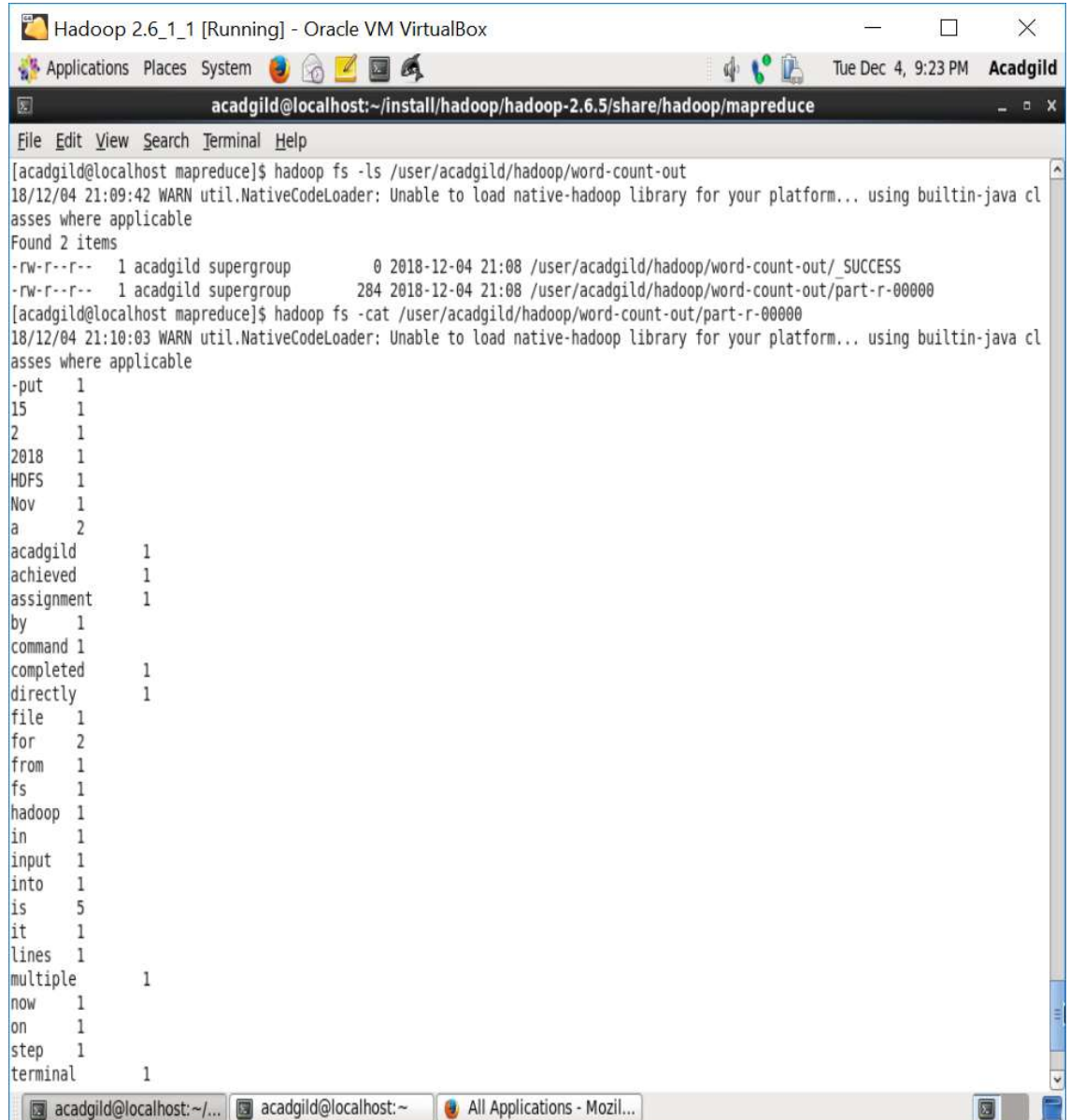
Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Memory Used	Memory Total	Memory Reserved	VCores Used	VCores Total	VCores Re
1	0	1	0	1	2 GB	8 GB	0 B	1	8	0

Show 20 entries

ID	User	Name	Application Type	Queue	StartTime	FinishTime	Stat
application_1543937045769_0001	acagild	word count	MAPREDUCE	default	Tue, 04 Dec 2018 15:38:23 GMT	N/A	ACCE

Showing 1 to 1 of 1 entries

- 8) Since only one reducer was triggered for processing this file, there was only one part-r\* file created in the output folder specified above. Once the job completed, the content of output file created by this job in /user/acadgild/hadoop/word-count-out/part-r-00000 was displayed using -cat option of Hadoop fs command. This file contained all the words in my input along with the count of occurrence of each of them.



```
Hadoop 2.6.1_1 [Running] - Oracle VM VirtualBox
Applications Places System
acadmild@localhost:~/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce
File Edit View Search Terminal Help
[acadmild@localhost mapreduce]$ hadoop fs -ls /user/acadmild/hadoop/word-count-out
18/12/04 21:09:42 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 acadmild supergroup 0 2018-12-04 21:08 /user/acadmild/hadoop/word-count-out/_SUCCESS
-rw-r--r-- 1 acadmild supergroup 284 2018-12-04 21:08 /user/acadmild/hadoop/word-count-out/part-r-00000
[acadmild@localhost mapreduce]$ hadoop fs -cat /user/acadmild/hadoop/word-count-out/part-r-00000
18/12/04 21:10:03 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
-put 1
15 1
2 1
2018 1
HDFS 1
Nov 1
a 2
acadmild 1
achieved 1
assignment 1
by 1
command 1
completed 1
directly 1
file 1
for 2
from 1
fs 1
hadoop 1
in 1
input 1
into 1
is 5
it 1
lines 1
multiple 1
now 1
on 1
step 1
terminal 1
```

## 9) Wordmean

Ran "hadoop jar hadoop-mapreduce-examples-2.6.5.jar wordmean /user/acadgild/hadoop/word-count.txt /user/acadgild/hadoop/wordmean"

```
Hadoop 2.6.1_1 [Running] - Oracle VM VirtualBox
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 jar hadoop-mapreduce-examples-2.6.5.jar wordmean /user/acadgild/hadoop/word-count.txt /user/acadgild/hadoop/wordmean
18/12/04 21:29:33 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/12/04 21:29:34 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/12/04 21:29:36 INFO input.FileInputFormat: Total input paths to process : 1
18/12/04 21:29:36 INFO mapreduce.JobSubmitter: number of splits:1
18/12/04 21:29:37 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1543937045769_0003
18/12/04 21:29:37 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1543937045769_0003/
18/12/04 21:29:37 INFO mapreduce.Job: Running job: job_1543937045769_0003
18/12/04 21:29:48 INFO mapreduce.Job: Job job_1543937045769_0003 running in uber mode : false
18/12/04 21:29:48 INFO mapreduce.Job: map 0% reduce 0%
18/12/04 21:29:55 INFO mapreduce.Job: map 100% reduce 0%
18/12/04 21:29:55 INFO mapreduce.Job: map 100% reduce 100%
18/12/04 21:30:07 INFO mapreduce.Job: Job job_1543937045769_0003 completed successfully
18/12/04 21:30:07 INFO mapreduce.Job: Counters: 49
File System Counters
FILE: Number of bytes read=39
FILE: Number of bytes written=215479
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=369
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)=5025
Total time spent by all map tasks (ms)=5025
Total time spent by all reduce tasks (ms)=8047
Total vcore-milliseconds taken by all map tasks=5025
Total vcore-milliseconds taken by all reduce tasks=8047
acadgild@localhost:~/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce
```

Hadoop 2.6.1\_1 [Running] - Oracle VM VirtualBox

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

All Applications - Mozilla Firefox

localhost:8088/cluster

Most Visited Getting Started yam

**hadoop**

**All Applications**

Cluster

Cluster Metrics

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Memory Used	Memory Total	Memory Reserved	Vcores Used	Vcores Total
3	0	0	3	0	0 B	8 GB	0 B	0	8

Show 20 entries

ID	User	Name	Application Type	Queue	StartTime	FinishTime
application_1543937045769_0003	acadgild	word mean	MAPREDUCE	default	Tue, 04 Dec 2018 15:59:37 GMT	Tue, 04 Dec 2018 16:00:05 GMT
application_1543937045769_0002	acadgild	word mean	MAPREDUCE	default	Tue, 04 Dec 2018 15:55:50 GMT	Tue, 04 Dec 2018 15:56:17 GMT
application_1543937045769_0001	acadgild	word count	MAPREDUCE	default	Tue, 04 Dec 2018 15:38:23 GMT	Tue, 04 Dec 2018 15:38:57 GMT

Showing 1 to 3 of 3 entries

acacgild@localhost:~/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce

```
Hadoop 2.6.1_1 [Running] - Oracle VM VirtualBox
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 ~]$ hadoop fs -ls /user/acadgild/hadoop/wordmean
18/12/04 21:31:07 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 2018-12-04 21:30 /user/acadgild/hadoop/wordmean/ SUCCESS
-rw-r--r-- 1 acadgild supergroup 20 2018-12-04 21:30 /user/acadgild/hadoop/wordmean/part-r-000000
[acadgild@localhost ~]$ hadoop fs -cat /user/acadgild/hadoop/wordmean/part-r-000000
18/12/04 21:31:25 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
count 48
length 198
[acadgild@localhost ~]$
```



## 10) Wordmedian

Ran "hadoop jar hadoop-mapreduce-examples-2.6.5.jar wordmedian /user/acadgild/hadoop/word-count.txt /user/acadgild/hadoop/wordmedian"

**Screenshot 1: Terminal Output**

```
acacgild@localhost:~/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce
You have new mail in /var/spool/mail/acacgild
[acacgild@localhost mapreduce]$ hadoop jar hadoop-mapreduce-examples-2.6.5.jar wordmedian /user/acacgild/hadoop/word-count.txt /user/acacgild/hadoop/wordmedian
18/12/04 21:33:11 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/12/04 21:33:12 INFO client.RMProxy: Connecting to ResourceManager at localhost:127.0.0.1:8032
18/12/04 21:33:13 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool in interface and execute your application with ToolRunner to remedy this.
18/12/04 21:33:14 INFO input.FileInputFormat: Total input paths to process : 1
18/12/04 21:33:14 INFO mapreduce.JobSubmitter: number of splits:1
18/12/04 21:33:14 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1543937045769_0004
18/12/04 21:33:14 INFO impl.YarnClientImpl: Submitted application application_1543937045769_0004
18/12/04 21:33:15 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1543937045769_0004/
18/12/04 21:33:15 INFO mapreduce.Job: Running job: job_1543937045769_0004
18/12/04 21:33:24 INFO mapreduce.Job: Job job_1543937045769_0004 running in uber mode : false
18/12/04 21:33:24 INFO mapreduce.Job: map 0% reduce 0%
18/12/04 21:33:31 INFO mapreduce.Job: map 100% reduce 0%
18/12/04 21:33:39 INFO mapreduce.Job: map 100% reduce 100%
18/12/04 21:33:40 INFO mapreduce.Job: Job job_1543937045769_0004 completed successfully
18/12/04 21:33:40 INFO mapreduce.Job: Counters: 49
File System Counters
FILE: Number of bytes read=100
FILE: Number of bytes written=215353
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=369
HDFS: Number of bytes written=43
HDFS: Number of read operations=0
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)=4781
Total time spent by all reduces in occupied slots (ms)=5601
Total time spent by all map tasks (ms)=4781
Total time spent by all reduce tasks (ms)=5601
```

**Screenshot 2: Hadoop Web Interface**

**All Applications**

Cluster Metrics									
Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Memory Used	Memory Total	Memory Reserved	Vcores Used	Vcores Total
4	0	0	4	0	0 B	8 GB	0 B	0	8

ID	User	Name	Application Type	Queue	StartTime	FinishTime
application_1543937045769_0004	acacgild	word median	MAPREDUCE	default	Tue, 04 Dec 2018 16:03:14 GMT	Tue, 04 Dec 2018 16:03:38 GMT
application_1543937045769_0003	acacgild	word mean	MAPREDUCE	default	Tue, 04 Dec 2018 15:59:37 GMT	Tue, 04 Dec 2018 16:00:05 GMT
application_1543937045769_0002	acacgild	word mean	MAPREDUCE	default	Tue, 04 Dec 2018 15:55:50 GMT	Tue, 04 Dec 2018 15:56:17 GMT
application_1543937045769_0001	acacgild	word count	MAPREDUCE	default	Tue, 04 Dec 2018 15:38:23 GMT	Tue, 04 Dec 2018 15:38:57 GMT

**Screenshot 3: Terminal Output**

```
acacgild@localhost:~/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce
You have new mail in /var/spool/mail/acacgild
[acacgild@localhost ~]$ hadoop fs -ls /user/acacgild/hadoop/wordmedian
18/12/04 21:34:38 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 acacgild supergroup 0 2018-12-04 21:33 /user/acacgild/hadoop/wordmedian/ SUCCESS
-rw-r--r-- 1 acacgild supergroup 43 2018-12-04 21:33 /user/acacgild/hadoop/wordmedian/part-r-000000
[acacgild@localhost ~]$ hadoop fs -cat /user/acacgild/hadoop/wordmedian/part-r-000000
1
2
3
4
5
6
7
8
9
10
```

## 11) Wordstandarddeviation

Ran "hadoop jar hadoop-mapreduce-examples-2.6.5.jar wordstandarddeviation /user/acadgild/hadoop/word-count.txt /user/acadgild/hadoop/wordstandarddeviation"

The first screenshot shows a terminal window in Oracle VM VirtualBox running Hadoop 2.6.1\_1. The command executed is `hadoop jar hadoop-mapreduce-examples-2.6.5.jar wordstandarddeviation /user/acadgild/hadoop/word-count.txt /user/acadgild/hadoop/wordstandarddeviation`. The output shows the job running successfully with 1 map task and 1 reduce task. The job ID is `application_1543937045769_0005`.

The second screenshot shows the Hadoop web interface in Mozilla Firefox. The "All Applications" page displays a table of running jobs. The job `application_1543937045769_0005` is shown as running.

App ID	User	Name	Application Type	Queue	Start Time	Finish Time
application_1543937045769_0005	acadgild	word stddev	MAPREDUCE	default	Tue, 04 Dec 2018 16:06:25 GMT	Tue, 04 Dec 2018 16:06:51 GMT
application_1543937045769_0004	acadgild	word median	MAPREDUCE	default	Tue, 04 Dec 2018 16:03:14 GMT	Tue, 04 Dec 2018 16:03:38 GMT
application_1543937045769_0003	acadgild	word mean	MAPREDUCE	default	Tue, 04 Dec 2018 15:59:37 GMT	Tue, 04 Dec 2018 16:00:05 GMT
application_1543937045769_0002	acadgild	word mean	MAPREDUCE	default	Tue, 04 Dec 2018 15:55:50 GMT	Tue, 04 Dec 2018 15:56:17 GMT

The third screenshot shows the terminal window again, displaying the output of the `hadoop fs -ls` command. It shows the file `/user/acadgild/hadoop/wordstandarddeviation/part-r-000000` has been created with a size of 48 bytes.