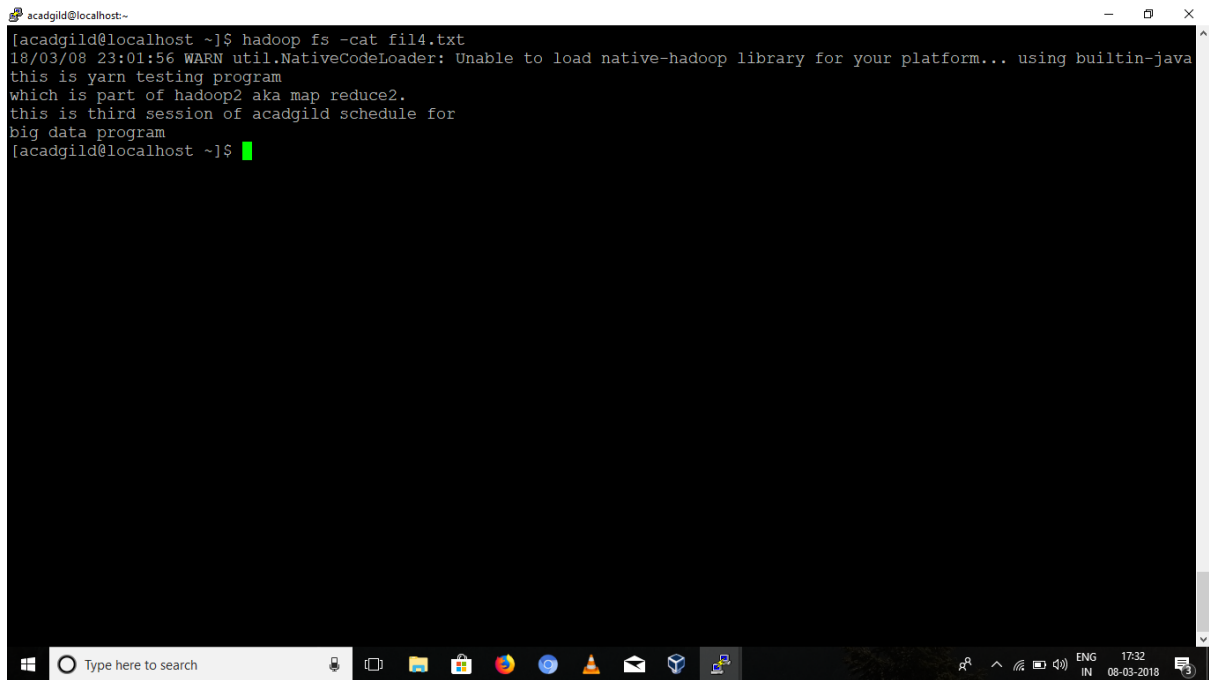


ASSIGNMENT 3.1

Task 1:

Command used:

- a. Cat > fil4.txt : creating a file in local file system
- b. Hadoop fs -put fil4.txt : pushing the file into hadoop file system
- c. Hadoop fs -cat fil4.txt : recheck the content of the file which is recently uploaded



```
acadgild@localhost:~$ hadoop fs -cat fil4.txt
18/03/08 23:01:56 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java
this is yarn testing program
which is part of hadoop2 aka map reduce2.
this is third session of acadgild schedule for
big data program
acadgild@localhost ~]$
```

- d. Execute **wordstandarddeviation** for Yarn:
 - i. Hadoop jar hadoop-mapreduce-examples-2.9.0.jar wordstandarddeviation fil4.txt wsd1
 - ii. Hadoop-mapreduce-examples-2.9.0.jar is the jar file
 - iii. Wordstandarddeviation is the key word to calculate standard deviation

```
acadgild@localhost:~/install/hadoop/hadoop-2.9.0/share/hadoop/mapreduce
[acadgild@localhost mapreduce]$ hadoop jar hadoop-mapreduce-examples-2.9.0.jar wordstandarddeviation fil4.txt wsd1
18/03/08 23:03:00 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java
classes where applicable
18/03/08 23:03:02 INFO client.RMPProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/03/08 23:03:04 INFO input.FileInputFormat: Total input files to process : 1
18/03/08 23:03:04 WARN hdfs.DataStreamer: Caught exception
java.lang.InterruptedException
    at java.lang.Object.wait(Native Method)
    at java.lang.Thread.join(Thread.java:1252)
    at java.lang.Thread.join(Thread.java:1326)
    at org.apache.hadoop.hdfs.DataStreamer.closeResponder(DataStreamer.java:980)
    at org.apache.hadoop.hdfs.DataStreamer.endBlock(DataStreamer.java:630)
    at org.apache.hadoop.hdfs.DataStreamer.run(DataStreamer.java:807)
18/03/08 23:03:04 WARN hdfs.DataStreamer: Caught exception
java.lang.InterruptedException
    at java.lang.Object.wait(Native Method)
    at java.lang.Thread.join(Thread.java:1252)
    at java.lang.Thread.join(Thread.java:1326)
    at org.apache.hadoop.hdfs.DataStreamer.closeResponder(DataStreamer.java:980)
    at org.apache.hadoop.hdfs.DataStreamer.endBlock(DataStreamer.java:630)
    at org.apache.hadoop.hdfs.DataStreamer.run(DataStreamer.java:807)
18/03/08 23:03:04 INFO mapreduce.JobSubmitter: number of splits:1
18/03/08 23:03:05 INFO Configuration.deprecation: yarn.resourcemanager.system-metrics-publisher.enabled is deprecated. Ins
tead, use yarn.system-metrics-publisher.enabled
18/03/08 23:03:05 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1520520168165_0010
18/03/08 23:03:06 INFO impl.YarnClientImpl: Submitted application application_1520520168165_0010
18/03/08 23:03:06 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1520520168165_0010
/
18/03/08 23:03:06 INFO mapreduce.Job: Running job: job_1520520168165_0010
```

```
acadgild@localhost:~/install/hadoop/hadoop-2.9.0/share/hadoop/mapreduce
Map input records=4
Map output records=72
Map output bytes=1056
Map output materialized bytes=56
Input split bytes=109
Combine input records=72
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)=159
CPU time spent (ms)=1460
Physical memory (bytes) snapshot=346943488
Virtual memory (bytes) snapshot=4127256576
Total committed heap usage (bytes)=222429184

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=136
File Output Format Counters
    Bytes Written=31
The standard deviation is: 2.1372197672053597
[acadgild@localhost mapreduce]$
```

- iv. Wsd1 is file created to store output
- v. `hadoop fs -ls wsd1` : to check the output reducer details
- vi. `hadoop fs -cat wsd1/part-r-00000` : to get the output details as show below.

```
acadmild@localhost:~$ hadoop fs -cat fil4.txt
18/03/08 23:01:56 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java
this is yarn testing program
which is part of hadoop2 aka map reduce2.
this is third session of acadmild schedule for
big data program
[acadmild@localhost ~]$ hadoop fs -ls wsd1
18/03/08 23:04:16 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 acadmild supergroup 0 2018-03-08 23:03 wsd1/_SUCCESS
-rw-r--r-- 1 acadmild supergroup 31 2018-03-08 23:03 wsd1/part-r-00000
[acadmild@localhost ~]$ hadoop fs -cat wsd1/part-r-00000
18/03/08 23:04:41 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java
classes where applicable
count 24
length 111
square 623
[acadmild@localhost ~]$
```

Word Mean execution:

Command used:

- Hadoop jar hadoop-mapreduce-examples-2.9.0.jar wordmean fil4.txt wmean
- hadoop-mapreduce-examples-2.9.0.jar is the jar file given.
- Wordmean is the keyword.
- Wmean is the file for output storage.

```
acadmild@localhost:~/install/hadoop/hadoop-2.9.0/share/hadoop/mapreduce$ hadoop jar hadoop-mapreduce-examples-2.9.0.jar wordmean fil4.txt wmean
18/03/08 23:11:36 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java
classes where applicable
18/03/08 23:11:38 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/03/08 23:11:40 INFO input.FileInputFormat: Total input files to process : 1
18/03/08 23:11:40 WARN hdfs.DataStreamer: Caught exception
java.lang.InterruptedExcepcion
    at java.lang.Object.wait(Native Method)
    at java.lang.Thread.join(Thread.java:1252)
    at java.lang.Thread.join(Thread.java:1326)
    at org.apache.hadoop.hdfs.DataStreamer.closeResponder(DataStreamer.java:980)
    at org.apache.hadoop.hdfs.DataStreamer.endBlock(DataStreamer.java:630)
    at org.apache.hadoop.hdfs.DataStreamer.run(DataStreamer.java:807)
18/03/08 23:11:40 WARN hdfs.DataStreamer: Caught exception
java.lang.InterruptedExcepcion
    at java.lang.Object.wait(Native Method)
    at java.lang.Thread.join(Thread.java:1252)
    at java.lang.Thread.join(Thread.java:1326)
    at org.apache.hadoop.hdfs.DataStreamer.closeResponder(DataStreamer.java:980)
    at org.apache.hadoop.hdfs.DataStreamer.endBlock(DataStreamer.java:630)
    at org.apache.hadoop.hdfs.DataStreamer.run(DataStreamer.java:807)
18/03/08 23:11:40 INFO mapreduce.JobSubmitter: number of splits:1
18/03/08 23:11:40 INFO Configuration.deprecation: yarn.resourcemanager.system-metrics-publisher.enabled is deprecated. Ins
tead, use yarn.system-metrics-publisher.enabled
18/03/08 23:11:41 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1520520168165_0013
18/03/08 23:11:41 INFO impl.YarnClientImpl: Submitted application application_1520520168165_0013
18/03/08 23:11:41 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1520520168165_0013
/
18/03/08 23:11:41 INFO mapreduce.Job: Running job: job_1520520168165_0013
```

```
acadgild@localhost:~/install/hadoop/hadoop-2.9.0/share/hadoop/mapreduce
Map input records=4
Map output records=48
Map output bytes=696
Map output materialized bytes=39
Input split bytes=109
Combine input records=48
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)=187
CPU time spent (ms)=1860
Physical memory (bytes) snapshot=344858624
Virtual memory (bytes) snapshot=4127256576
Total committed heap usage (bytes)=222429184

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=136
File Output Format Counters
Bytes Written=20

The mean is: 4.625
[acadgild@localhost mapreduce]$
```

- e. Hadoop fs -ls wmean : to get the output details of the reducer used
- f. Hadoop fs -cat wmean/part-r-00000

```
acadgild@localhost:~$ hadoop fs -ls wmean
18/03/08 23:12:28 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 2018-03-08 23:12 wmean/ SUCCESS
-rw-r--r-- 1 acadgild supergroup 20 2018-03-08 23:12 wmean/part-r-00000
[acadgild@localhost ~]$ hadoop fs -cat wmean/part-r-00000
18/03/08 23:12:42 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java
classes where applicable
count 24
length 111
[acadgild@localhost ~]$
```

Word Median

Command used:

- Hadoop jar hadoop-mapreduce-examples-2.9.0.jar wordmedian fil4.txt wm
- hadoop-mapreduce-examples-2.9.0.jar is the given jar file
- wordmedian is the keyword to execute word median for yarn.
- Wm is the file for output storage.

```
acacgild@localhost:~/install/hadoop/hadoop-2.9.0/share/hadoop/mapreduce
[acacgild@localhost mapreduce]$ hadoop jar hadoop-mapreduce-examples-2.9.0.jar wordmedian fil4.txt wm
18/03/08 23:08:44 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java
classes where applicable
18/03/08 23:08:46 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/03/08 23:08:48 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool
interface and execute your application with ToolRunner to remedy this.
18/03/08 23:08:48 INFO input.FileInputFormat: Total input files to process : 1
18/03/08 23:08:49 WARN hdfs.DataStreamer: Caught exception
java.lang.InterruptedExce
at java.lang.Object.wait(Native Method)
at java.lang.Thread.join(Thread.java:1252)
at java.lang.Thread.join(Thread.java:1326)
at org.apache.hadoop.hdfs.DataStreamer.closeResponder(DataStreamer.java:980)
at org.apache.hadoop.hdfs.DataStreamer.endBlock(DataStreamer.java:630)
at org.apache.hadoop.hdfs.DataStreamer.run(DataStreamer.java:807)
18/03/08 23:08:49 WARN hdfs.DataStreamer: Caught exception
java.lang.InterruptedExce
at java.lang.Object.wait(Native Method)
at java.lang.Thread.join(Thread.java:1252)
at java.lang.Thread.join(Thread.java:1326)
at org.apache.hadoop.hdfs.DataStreamer.closeResponder(DataStreamer.java:980)
at org.apache.hadoop.hdfs.DataStreamer.endBlock(DataStreamer.java:630)
at org.apache.hadoop.hdfs.DataStreamer.run(DataStreamer.java:807)
18/03/08 23:08:49 INFO mapreduce.JobSubmitter: number of splits:1
18/03/08 23:08:49 INFO Configuration.deprecation: yarn.resourcemanager.system-metrics-publisher.enabled is deprecated. Ins
tead, use yarn.system-metrics-publisher.enabled
18/03/08 23:08:49 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1520520168165_0012
18/03/08 23:08:50 INFO impl.YarnClientImpl: Submitted application application_1520520168165_0012
18/03/08 23:08:50 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1520520168165_0012
/
18/03/08 23:08:50 INFO mapreduce.Job: Running job: job_1520520168165_0012
```

```
acacgild@localhost:~/install/hadoop/hadoop-2.9.0/share/hadoop/mapreduce
Map input records=4
Map output records=24
Map output bytes=192
Map output materialized bytes=66
Input split bytes=109
Combine input records=24
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)=152
CPU time spent (ms)=1650
Physical memory (bytes) snapshot=347500544
Virtual memory (bytes) snapshot=4127256576
Total committed heap usage (bytes)=222429184

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=136
File Output Format Counters
Bytes Written=24

The median is: 4
[acacgild@localhost mapreduce]$
```

- e. Hadoop fs -ls wm : to get the output details of reducer used
- f. Hadoop fs -cat wm/part-r-00000 : to get the output

```
acagild@localhost:~$ hadoop fs -ls wm
18/03/08 23:10:34 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 acagild supergroup          0 2018-03-08 23:09 wm/ SUCCESS
-rw-r--r-- 1 acagild supergroup        24 2018-03-08 23:09 wm/part-r-00000
[acagild@localhost ~]$ hadoop fs -cat wm/part-r-00000
18/03/08 23:10:50 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java
classes where applicable
2      5
3      4
4      5
5      2
7      5
8      3
[acagild@localhost ~]$
```