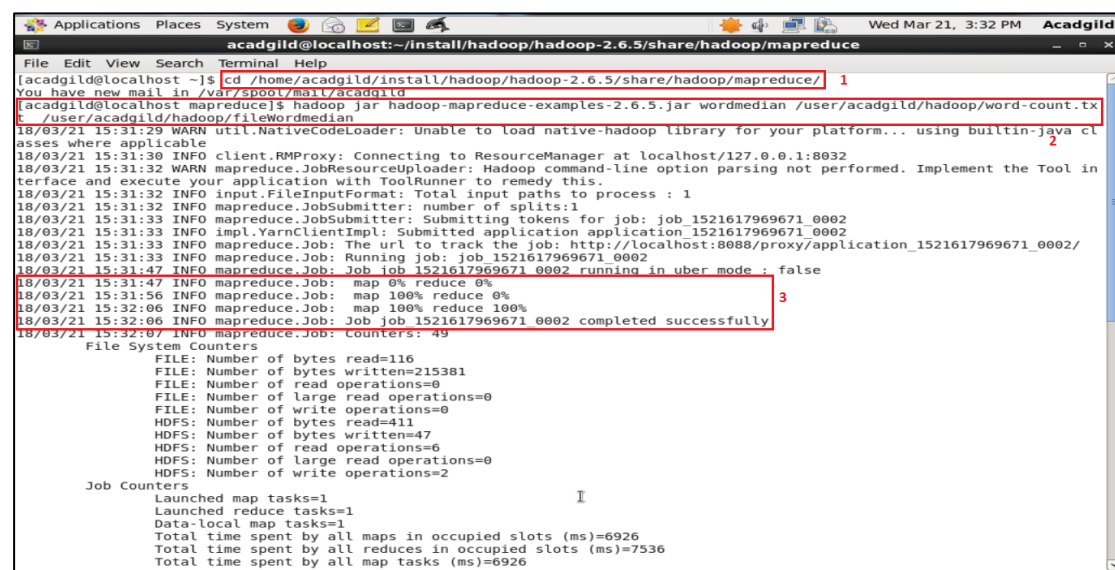


Session 3 - YARN

Assignment 1

Task1 : Execute **WordMedian** , **WordMean** , **WordStandardDeviation** programs using **hadoop-mapreduce-examples-2.9.0.jar** file present in your AcadGild VM.

WordMedian :



```
acadgild@localhost:~/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce
[acadgild@localhost ~]$ cd /home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce/
[acadgild@localhost mapreduce]$ hadoop jar hadoop-mapreduce-examples-2.6.5.jar wordmedian /user/acadgild/hadoop/word-count.txt /user/acadgild/hadoop/fileWordmedian
18/03/21 15:31:29 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/03/21 15:31:30 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/03/21 15:31:32 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
18/03/21 15:31:32 INFO input.FileInputFormat: Total input paths to process : 1
18/03/21 15:31:32 INFO mapreduce.JobSubmitter: number of splits:1
18/03/21 15:31:33 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1521617969671_0002
18/03/21 15:31:33 INFO impl.YarnClientImpl: Submitted application application_1521617969671_0002
18/03/21 15:31:33 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1521617969671_0002/
18/03/21 15:31:33 INFO mapreduce.Job: Running job: job_1521617969671_0002
18/03/21 15:31:47 INFO mapreduce.Job: Job job 1521617969671_0002 running in uber mode : false
18/03/21 15:31:56 INFO mapreduce.Job: map 100% reduce 0%
18/03/21 15:32:06 INFO mapreduce.Job: map 100% reduce 100%
18/03/21 15:32:06 INFO mapreduce.Job: Job job 1521617969671_0002 completed successfully
18/03/21 15:32:07 INFO mapreduce.Job: Counters: 49
  File System Counters
    FILE: Number of bytes read=116
    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=411
    HDFS: Number of bytes written=47
    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)=6926
    Total time spent by all reduces in occupied slots (ms)=7536
    Total time spent by all map tasks (ms)=6926
```

With reference to the screenshot above,

1 : Move to the folder where the jar file is located using the following command :

```
cd /home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce/
```

2 : Execute the WordMedian class from the jar file by passing 2 parameter to it : the source location of the file and a location for the target files to be stored. So the command will be :

```
hadoop jar hadoop-mareduce-examples-2.6.5.jar wordmedian
/user/acadgild/hadoop/word-count.txt /user/acadgild/hadoop/fileWordmedian
```

3 : As the jar is executed a map reduce job is initiated. The mapper is executed completely first and then the reducer job is picked up. Once both are executed, the job is said to be completed successfully.

```
Applications Places System acadgild@localhost:~/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce
File Edit View Search Terminal Help
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=289
File Output Format Counters
Bytes Written=47
The median is: 5
[acadgild@localhost mapreduce]$ hdfs dfs -ls /user/acadgild/hadoop/fileWordmedian
18/03/21 15:34:58 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 acadgild supergroup 0 2018-03-21 15:32 /user/acadgild/hadoop/fileWordmedian/ SUCCESS
-rw-r--r-- 1 acadgild supergroup 47 2018-03-21 15:32 /user/acadgild/hadoop/fileWordmedian/part-r-000000
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost mapreduce]$ hdfs dfs -cat /user/acadgild/hadoop/fileWordmedian/part-r-000000
18/03/21 15:35:54 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
cat: /user/acadgild/hadoop/fileWordmedian/part-r-000000: No such file or directory
[acadgild@localhost mapreduce]$ hdfs dfs -cat /user/acadgild/hadoop/fileWordmedian/part-r-000000
18/03/21 15:36:46 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
2 8
3 8
4 3
5 4
6 3
7 5
8 4
9 5
10 2
11 1
12 1
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost mapreduce]$
```

With reference to the screenshot above,

1 : Once the job is completed, the output is printed as follows :

The median is: 5

2 : When navigated to the location given in the hadoop jar command, there are 2 files that are created. The command used is

hdfs dfs -ls /user/acadgild/hadoop/fileWordmedian

3 : `_SUCCESS` : indicating that the job is completed successfully.

`part-r-000000` : the part file (a partition) that contains the output of the job.

4 : To view the contents of the part file, a cat command is used.

hdfs dfs -cat /user/acadgild/hadoop/fileWordmedian/part-r-000000

WordMean :

```

[acadgild@localhost mapreduce]$ hadoop jar hadoop-mapreduce-examples-2.6.5.jar wordmean /user/acadgild/hadoop/word-count.txt /user/acadgild/hadoop/fileWordmean
18/03/21 15:40:45 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/03/21 15:40:46 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/03/21 15:40:48 INFO input.FileInputFormat: Total input paths to process : 1
18/03/21 15:40:49 INFO mapreduce.JobSubmitter: number of splits:1
18/03/21 15:40:49 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1521617969671_0003
18/03/21 15:40:49 INFO impl.YarnClientImpl: Submitted application application_1521617969671_0003
18/03/21 15:40:50 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1521617969671_0003/
18/03/21 15:40:50 INFO mapreduce.Job: Running job: job_1521617969671_0003
18/03/21 15:41:01 INFO mapreduce.Job: Job job_1521617969671_0003 running in uber mode : false
18/03/21 15:41:01 INFO mapreduce.Job: map 0% reduce 0%
18/03/21 15:41:10 INFO mapreduce.Job: map 100% reduce 0%
18/03/21 15:41:19 INFO mapreduce.Job: map 100% reduce 100%
18/03/21 15:41:19 INFO mapreduce.Job: Job job_1521617969671_0003 completed successfully
18/03/21 15:41:19 INFO mapreduce.Job: Counters: 49
File System Counters
  FILE: Number of bytes read=39
  FILE: Number of bytes written=215487
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=411
  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)=6274
  Total time spent by all reduces in occupied slots (ms)=6841
  Total time spent by all map tasks (ms)=6274
  Total time spent by all reduce tasks (ms)=6841
  Total vcore-milliseconds taken by all map tasks=6274
  Total vcore-milliseconds taken by all reduce tasks=6841
  Total megabyte-milliseconds taken by all map tasks=6424576

```

With reference to the screenshot above,

1 : Execute the WordMean class from the jar file by passing 2 parameter to it : the source location of the file and a location for the target files to be stored. So the command will be :

*hadoop jar hadoop-mareduce-examples-2.6.5.jar wordmean
/user/acadgild/hadoop/word-count.txt /user/acadgild/hadoop/fileWordmean*

2 : As the jar is executed a map reduce job is initiated. The mapper is executed completely first and then the reducer job is picked up. Once both are executed, the job is said to be completed successfully.

```

[acadgild@localhost mapreduce]$ hdfs dfs -ls /user/acadgild/hadoop/fileWordmean
18/03/21 15:43:41 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 2018-03-21 15:41 /user/acadgild/hadoop/fileWordmean/ SUCCESS
-rw-r--r-- 1 acadgild supergroup 20 2018-03-21 15:41 /user/acadgild/hadoop/fileWordmean/part-r-00000
[acadgild@localhost mapreduce]$ hdfs dfs -cat /user/acadgild/hadoop/fileWordmean/part-r-00000
count 44
length 245
[acadgild@localhost mapreduce]$

```

With reference to the screenshot above,

1 : Once the job is completed, the output is printed as follows :

The mean is: 5.5681818181818181

2 : When navigated to the location given in the hadoop jar command, there are 2 files that are created. The command used is

hdfs dfs -ls /user/acadgild/hadoop/fileWordmean

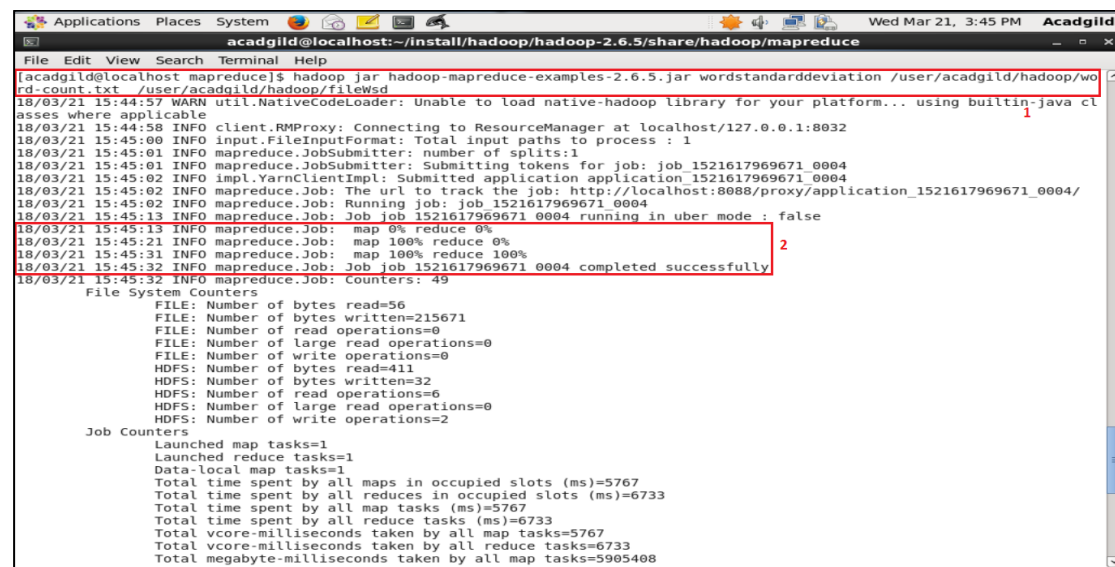
3 : `_SUCCESS` : indicating that the job is completed successfully.

`part-r-00000` : the part file (a partition) that contains the output of the job.

4 : To view the contents of the part file, a `cat` command is used.

hdfs dfs -cat /user/acadgild/hadoop/fileWordmean/part-r-00000

WordStandardDeviation :



```
acadgild@localhost:~/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce
File Edit View Search Terminal Help
[acadgild@localhost mapreduce]$ hadoop jar hadoop-mapreduce-examples-2.6.5.jar wordstandarddeviation /user/acadgild/hadoop/wo
rd-count.txt /user/acadgild/hadoop/fileWsd
18/03/21 15:44:57 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
18/03/21 15:44:58 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/03/21 15:45:00 INFO input.FileInputFormat: Total input paths to process : 1
18/03/21 15:45:01 INFO mapreduce.JobSubmitter: number of splits:1
18/03/21 15:45:01 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1521617969671_0004
18/03/21 15:45:02 INFO impl.YarnClientImpl: Submitted application application_1521617969671_0004
18/03/21 15:45:02 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1521617969671_0004/
18/03/21 15:45:02 INFO mapreduce.Job: Running job: job_1521617969671_0004
18/03/21 15:45:13 INFO mapreduce.Job: Job job_1521617969671_0004 running in uber mode : false
18/03/21 15:45:13 INFO mapreduce.Job: map 0% reduce 0%
18/03/21 15:45:21 INFO mapreduce.Job: map 100% reduce 0%
18/03/21 15:45:31 INFO mapreduce.Job: map 100% reduce 100%
18/03/21 15:45:32 INFO mapreduce.Job: Job job_1521617969671_0004 completed successfully
18/03/21 15:45:32 INFO mapreduce.Job: Counters: 49
  File System Counters
    FILE: Number of bytes read=56
    FILE: Number of bytes written=215671
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=411
    HDFS: Number of bytes written=32
    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)=5767
    Total time spent by all reduces in occupied slots (ms)=6733
    Total time spent by all map tasks (ms)=5767
    Total time spent by all reduce tasks (ms)=6733
    Total vcore-milliseconds taken by all map tasks=5767
    Total vcore-milliseconds taken by all reduce tasks=6733
    Total megabyte-milliseconds taken by all map tasks=5905408
```

With reference to the screenshot above,

1 : Execute the `WordStandardDeviation` class from the jar file by passing 2 parameter to it : the source location of the file and a location for the target files to be stored. So the command will be :

*hadoop jar hadoop-mareduce-examples-2.6.5.jar wordstandarddeviation
/user/acadgild/hadoop/word-count.txt /user/acadgild/hadoop/fileWsd*

2 : As the jar is executed a map reduce job is initiated. The mapper is executed completely first and then the reducer job is picked up. Once both are executed, the job is said to be completed successfully.

```
Acadgild@localhost:~/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce
File Edit View Search Terminal Help
Merged Map outputs=1
GC time elapsed (ms)=174
CPU time spent (ms)=1890
Physical memory (bytes) snapshot=297906176
Virtual memory (bytes) snapshot=4118237184
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=289
File Output Format Counters
Bytes Written=32
The standard deviation is: 2.8872582722331077 1
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost mapreduce]$ hdfs dfs -ls /user/acadgild/hadoop/fileWsd 2
18/03/21 15:46:17 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 acadgild supergroup 0 2018-03-21 15:45 /user/acadgild/hadoop/fileWsd/_SUCCESS 3
-rw-r--r-- 1 acadgild supergroup 32 2018-03-21 15:45 /user/acadgild/hadoop/fileWsd/part-r-000000 4
[acadgild@localhost mapreduce]$ hdfs dfs -cat /user/acadgild/hadoop/fileWsd/part-r-000000 5
18/03/21 15:46:32 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
count 44
length 245
square 1731
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost mapreduce]$
```

With reference to the screenshot above,

1 : Once the job is completed, the output is printed as follows :

The standard deviation is: 2.8872582722331077

2 : When navigated to the location given in the hadoop jar command, there are 2 files that are created. The command used is

hdfs dfs -ls /user/acadgild/hadoop/fileWsd

3 : `_SUCCESS` : indicating that the job is completed successfully.

`part-r-000000` : the part file (a partition) that contains the output of the job.

4 : To view the contents of the part file, a cat command is used.

hdfs dfs -cat /user/acadgild/hadoop/fileWsd/part-r-000000