

ASSIGNMENT YARN

Problem Statement

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

EXPLANATION: SO FIRST WE CREATED A TEXT FILE "Worddata.txt" IN WHICH ENTERED SOME WORDS. THEN WE COPIED THE SAME FILE INTO HDFS USING "put" COMMAND.

CODE: `hadoop fs -put Worddata.txt /user/acadgild`

SOLUTION REPORT:

```
[acadgild@localhost ~]$ cat Worddata.txt
he federal shutdown has ended, at least temporarily, in squalid defeat for
President Trump. He failed to secure funding for his signature border wall,
and none seems forthcoming. True to form, the prolonged shutdown eclipsed
another immigration story that Trump had pushed to the top of the news cycle
just a couple months earlier: the caravan of 6,000 or so Central American
migrants making their way toward the United States. They arrived in Tijuana
in November to find the border less hospitable than they had been led to
believe: a pre-existing backlog meant that they would not be granted speedy
&acirc;credible fear&acirc; hearings to start the asylum application process, with
waitlists that stretched in some cases for months. Some attempted to cross
the border anyway, some went home, but most waited, shuttling between
makeshift shelters in a city overwhelmed.
```

```
[acadgild@localhost ~]$ hadoop fs -put Worddata.txt /user/acadgild
19/03/20 13:50:32 WARN util.NativeCodeLoader: Unable to load native-hadoop
library for your platform... using builtin-java classes where applicable
```

EXPLANATION: NOW WE CREATED A DIRECTORY IN LOCAL "executables" IN WHICH CREATED THREE FOLDERS AS "wordMean", "wordMedian" AND "wordStandardDeviation". THEN COPIED THE JAR FILE "hadoop-mapreduce-example-2.6.5.jar" WHICH IS REQUIRED FOR MAPREDUCE PROGRAM TO RUN. THEN EXTRACTED THE "driver.class", "mapper.class" AND "reducer.class" FOR EACH INTO RESPECTIVE FOLDERS.

1) WORD MEAN:

EXPLANATION: NOW FOR THE EXAMPLE "WORD MEAN". WE WILL FIRST CHANGE THE DIRECTORY TO "executables" THEN AGAIN CHANGE TO "wordMean". THEN WE DO THE BELOW CODE TO CREATE A MANIFEST FOR "WORD MEAN" WHICH WILL CREATE A JAR FILE FOR THE SAME.

CODE: jar -cvf -C /home/acadgild/executables/wordMean

NOW NEXT, WE THEN ENTERED THE FOLLOWING CODE TO RUN OUR JAR FILE. IN WHICH PROVIDE INPUT FILE TO EXECUTE ON AND ENTER A OUTPUT PATH WHERE THE FILE BE SAVED IN HDFS:

CODE: hadoop jar hadoop-mapreduce-examples-2.6.5.jar wordmean /user/acadgild/Worddata.txt /user/acadgild/Hadoop

AND WE CAN CHECK THE OUPUT BY LOGGING INTO HDFS SHELL AND ENTERING THE BELOW COMMAND:

CODE: hadoop fs -cat /user/acadgild/Hadoop/part-r-00000

THE FILE "part-r-00000" IS THE FILE CREATED BY THE MAPREDUCE PROGRAM.

SOLUTION REPORT:

```
[acadgild@localhost ~]$ cd executables
You have new mail in /var/spool/mail/acadgild

[acadgild@localhost executables]$ ls
hadoop-mapreduce-examples-2.6.5.jar  wordMedian
wordMean                               wordStandardDeviation
You have new mail in /var/spool/mail/acadgild

[acadgild@localhost executables]$ cd wordMean

[acadgild@localhost wordMean]$ jar -cvf -C
/home/acadgild/executables/wordMean
added manifest
adding: home/acadgild/executables/wordMean/(in = 0) (out= 0) (stored 0%)
adding:
home/acadgild/executables/wordMean/WordMean$WordMeanMapper.class(in =
2459) (out= 978) (deflated 60%)
adding:
home/acadgild/executables/wordMean/WordMean$WordMeanReducer.class(in =
2459) (out= 923) (deflated 62%)
adding:
home/acadgild/executables/wordMean/hadoop-mapreduce-examples-2.6.5.jar(
in = 292710) (out= 269851) (deflated 7%)
adding: home/acadgild/executables/wordMean/WordMean.class(in = 5121)
(out= 2496) (deflated 51%)

[acadgild@localhost wordMean]$ hadoop jar
hadoop-mapreduce-examples-2.6.5.jar wordmean /user/acadgild/Worddata.txt
/user/acadgild/Hadoop
19/03/20 13:51:07 WARN util.NativeCodeLoader: Unable to load native-hadoop
library for your platform... using builtin-java classes where applicable
19/03/20 13:51:09 INFO client.RMProxy: Connecting to ResourceManager at
localhost/127.0.0.1:8032
19/03/20 13:51:11 INFO input.FileInputFormat: Total input paths to process
: 1
```

19/03/20 13:51:12 INFO mapreduce.JobSubmitter: number of splits:1
19/03/20 13:51:13 INFO mapreduce.JobSubmitter: Submitting tokens for job:
job_1553060650698_0008
19/03/20 13:51:14 INFO impl.YarnClientImpl: Submitted application
application_1553060650698_0008
19/03/20 13:51:15 INFO mapreduce.Job: The url to track the job:
http://localhost:8088/proxy/application_1553060650698_0008/
19/03/20 13:51:15 INFO mapreduce.Job: Running job: job_1553060650698_0008
19/03/20 13:51:37 INFO mapreduce.Job: Job job_1553060650698_0008 running
in uber mode : false
19/03/20 13:51:37 INFO mapreduce.Job: map 0% reduce 0%
19/03/20 13:51:52 INFO mapreduce.Job: map 100% reduce 0%
19/03/20 13:52:07 INFO mapreduce.Job: map 100% reduce 100%
19/03/20 13:52:08 INFO mapreduce.Job: Job job_1553060650698_0008 completed
successfully
19/03/20 13:52:09 INFO mapreduce.Job: Counters: 49

File System Counters

FILE: Number of bytes read=39
FILE: Number of bytes written=215451
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=980
HDFS: Number of bytes written=21
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)=11854
Total time spent by all reduces in occupied slots (ms)=12386
Total time spent by all map tasks (ms)=11854
Total time spent by all reduce tasks (ms)=12386
Total vcore-milliseconds taken by all map tasks=11854
Total vcore-milliseconds taken by all reduce tasks=12386
Total megabyte-milliseconds taken by all map tasks=12138496
Total megabyte-milliseconds taken by all reduce tasks=12683264

Map-Reduce Framework

Map input records=1
Map output records=278
Map output bytes=4031
Map output materialized bytes=39
Input split bytes=117
Combine input records=278
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)=330
CPU time spent (ms)=2920
Physical memory (bytes) snapshot=283222016
Virtual memory (bytes) snapshot=4118224896
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=863
File Output Format Counters
  Bytes Written=21
The mean is: 5.179856115107913
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost wordMean]$

[acadgild@localhost ~]$ hadoop fs -ls /user/acadgild
19/03/20 13:56:10 WARN util.NativeCodeLoader: Unable to load native-hadoop
library for your platform... using builtin-java classes where applicable
Found 6 items
drwxr-xr-x   - acadgild supergroup          0 2019-03-20 13:52
/user/acadgild/Hadoop
-rw-r--r--   1 acadgild supergroup          863 2019-03-20 13:50
/user/acadgild/Worddata.txt
-rw-r--r--   1 acadgild supergroup        2130 2019-03-20 13:18
/user/acadgild/data.txt
-rw-r--r--   1 acadgild supergroup    2283410 2019-03-20 12:31
/user/acadgild/movies.csv
drwxr-xr-x   - acadgild supergroup          0 2019-03-20 12:31
/user/acadgild/out
-rw-r--r--   1 acadgild supergroup          0 2019-03-20 12:31
/user/acadgild/ratings.csv
You have new mail in /var/spool/mail/acadgild

[acadgild@localhost ~]$ hadoop fs -ls /user/acadgild/Hadoop
19/03/20 13:56:25 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 2019-03-20 13:52
/user/acadgild/Hadoop/_SUCCESS
-rw-r--r--   1 acadgild supergroup         21 2019-03-20 13:52
/user/acadgild/Hadoop/part-r-00000

[acadgild@localhost ~]$ hadoop fs -cat /user/acadgild/Hadoop/part-r-00000
19/03/20 13:57:02 WARN util.NativeCodeLoader: Unable to load native-hadoop
library for your platform... using builtin-java classes where applicable
count 139

```

length 720

```
[acadgild@localhost ~]$ hadoop fs -cat /user/acadgild/Hadoop/part-r-00000
19/03/20 14:30:08 WARN util.NativeCodeLoader: Unable to load native-hadoop libra
ry for your platform... using builtin-java classes where applicable
count    139
length   720
[acadgild@localhost ~]$
```

So Word Mean = 5.179856115107913

2) WORD MEDIAN:

EXPLANATION: WE DO THE SAME PROCEDURE AS WE DID FOR "WORD MEAN". WE CARRY ON THE SAME STEPS N GET THE OUTPUT. THE MAPREDUCE OUTPUT FOR "WORD MEDIAN" IS SAVED IN HDFS DIRECTORY "Hadoop1" AND FILE "part-r-00000".

SOLUTION REPORT:

```
[acadgild@localhost Documents]$ cd executables
```

```
[acadgild@localhost executables]$ ls
hadoop-mapreduce-examples-2.6.5.jar    wordMedian
wordMean                                wordStandardDeviation
```

```
[acadgild@localhost executables]$ cd wordMedian
You have new mail in /var/spool/mail/acadgild
```

```
[acadgild@localhost wordMedian]$ jar -cvf -C
/home/acadgild/executables/wordMedian
added manifest
adding: home/acadgild/executables/wordMedian/(in = 0) (out= 0) (stored 0%)
adding:
home/acadgild/executables/wordMedian/WordMedian$WordMedianReducer.class
(in = 2454) (out= 890) (deflated 63%)
adding:
home/acadgild/executables/wordMedian/hadoop-mapreduce-examples-2.6.5.jar
r(in = 292710) (out= 269851) (deflated 7%)
adding: home/acadgild/executables/wordMedian/WordMedian.class(in = 5925)
(out= 2859) (deflated 51%)
adding:
home/acadgild/executables/wordMedian/WordMedian$WordMedianMapper.class(
in = 2385) (out= 932) (deflated 60%)
```

```
[acadgild@localhost wordMedian]$ hadoop jar
hadoop-mapreduce-examples-2.6.5.jar wordmedian
/user/acadgild/Worddata.txt /user/acadgild/Hadoop1
19/03/20 14:18:19 WARN util.NativeCodeLoader: Unable to load native-hadoop
library for your platform... using builtin-java classes where applicable
19/03/20 14:18:21 INFO client.RMProxy: Connecting to ResourceManager at
```

```
localhost/127.0.0.1:8032
19/03/20 14:18:23 WARN mapreduce.JobResourceUploader: Hadoop command-line
option parsing not performed. Implement the Tool interface and execute your
application with ToolRunner to remedy this.
19/03/20 14:18:24 INFO input.FileInputFormat: Total input paths to process
: 1
19/03/20 14:18:24 INFO mapreduce.JobSubmitter: number of splits:1
19/03/20 14:18:24 INFO mapreduce.JobSubmitter: Submitting tokens for job:
job_1553060650698_0009
19/03/20 14:18:26 INFO impl.YarnClientImpl: Submitted application
application_1553060650698_0009
19/03/20 14:18:26 INFO mapreduce.Job: The url to track the job:
http://localhost:8088/proxy/application_1553060650698_0009/
19/03/20 14:18:26 INFO mapreduce.Job: Running job: job_1553060650698_0009
19/03/20 14:18:46 INFO mapreduce.Job: Job job_1553060650698_0009 running
in uber mode : false
19/03/20 14:18:46 INFO mapreduce.Job:  map 0% reduce 0%
19/03/20 14:19:04 INFO mapreduce.Job:  map 100% reduce 0%
19/03/20 14:19:23 INFO mapreduce.Job:  map 100% reduce 100%
19/03/20 14:19:25 INFO mapreduce.Job: Job job_1553060650698_0009 completed
successfully
19/03/20 14:19:26 INFO mapreduce.Job: Counters: 49
  File System Counters
    FILE: Number of bytes read=126
    FILE: Number of bytes written=215363
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=980
    HDFS: Number of bytes written=58
    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)=16167
    Total time spent by all reduces in occupied slots (ms)=14185
    Total time spent by all map tasks (ms)=16167
    Total time spent by all reduce tasks (ms)=14185
    Total vcore-milliseconds taken by all map tasks=16167
    Total vcore-milliseconds taken by all reduce tasks=14185
    Total megabyte-milliseconds taken by all map tasks=16555008
    Total megabyte-milliseconds taken by all reduce tasks=14525440
  Map-Reduce Framework
    Map input records=1
    Map output records=139
    Map output bytes=1112
    Map output materialized bytes=126
    Input split bytes=117
    Combine input records=139
    Combine output records=12
```

```

        Reduce input groups=12
        Reduce shuffle bytes=126
        Reduce input records=12
        Reduce output records=12
        Spilled Records=24
        Shuffled Maps =1
        Failed Shuffles=0
        Merged Map outputs=1
        GC time elapsed (ms)=381
        CPU time spent (ms)=3000
        Physical memory (bytes) snapshot=284119040
        Virtual memory (bytes) snapshot=4118228992
        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=863
    File Output Format Counters
        Bytes Written=58
The median is: 5
You have new mail in /var/spool/mail/acadgild

[acadgild@localhost ~]$ hadoop fs -ls /user/acadgild/Hadoop1
19/03/20 14:21:09 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 2019-03-20 14:19
/user/acadgild/Hadoop1/_SUCCESS
-rw-r--r--    1 acadgild supergroup        58 2019-03-20 14:19
/user/acadgild/Hadoop1/part-r-00000
You have new mail in /var/spool/mail/acadgild

[acadgild@localhost ~]$ hadoop fs -cat
/user/acadgild/Hadoop1/part-r-00000
19/03/20 14:21:38 WARN util.NativeCodeLoader: Unable to load native-hadoop
library for your platform... using builtin-java classes where applicable
1      3
2      20
3      21
4      21
5      16
6      16
7      15
8      11
9      9
10     1
11     2
12     4

```

```
[acadgild@localhost ~]$ hadoop fs -cat /user/acadgild/Hadoop1/part-r-00000
19/03/20 14:21:38 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
1      3
2     20
3     21
4     21
5     16
6     16
7     15
8     11
9      9
10     1
11     2
12     4
[acadgild@localhost ~]$
```

So Word Mean is = 5

3) WORD STANDARD DEVIATION:

EXPLANATION: WE DO THE SAME PROCEDURE AS WE DID FOR "WORD MEAN". WE CARRY ON THE SAME STEPS N GET THE OUTPUT. THE MAPREDUCE OUTPUT FOR "WORD MEDIAN" IS SAVED IN HDFS DIRECTORY "Hadoop2" AND FILE "part-r-00000".

SOLUTION REPORT:

```
[acadgild@localhost ~]$ cd executables
```

```
[acadgild@localhost executables]$ cd wordStandardDeviation
```

```
[acadgild@localhost wordStandardDeviation]$ ls
hadoop-mapreduce-examples-2.6.5.jar
WordStandardDeviation.class
WordStandardDeviation$WordStandardDeviationMapper.class
WordStandardDeviation$WordStandardDeviationReducer.class
```

```
[acadgild@localhost wordStandardDeviation]$ jar -cvf -C
/home/acadgild/executables/wordStandardDeviation
added manifest
adding: home/acadgild/executables/wordStandardDeviation/(in = 0) (out=
0) (stored 0%)
adding:
home/acadgild/executables/wordStandardDeviation/WordStandardDeviation$W
ordStandardDeviationMapper.class(in = 2712) (out= 1069) (deflated 60%)
adding:
home/acadgild/executables/wordStandardDeviation/WordStandardDeviation$W
ordStandardDeviationReducer.class(in = 2549) (out= 936) (deflated 63%)
adding:
home/acadgild/executables/wordStandardDeviation/hadoop-mapreduce-exampl
es-2.6.5.jar(in = 292710) (out= 269851) (deflated 7%)
adding:
home/acadgild/executables/wordStandardDeviation/WordStandardDeviation.c
lass(in = 5660) (out= 2715) (deflated 52%)
```



```
[acadgild@localhost wordStandardDeviation]$ hadoop jar
hadoop-mapreduce-examples-2.6.5.jar wordstandarddeviation
/user/acadgild/Worddata.txt /user/acadgild/Hadoop2
19/03/20 14:26:40 WARN util.NativeCodeLoader: Unable to load native-hadoop
library for your platform... using builtin-java classes where applicable
19/03/20 14:26:42 INFO client.RMPProxy: Connecting to ResourceManager at
localhost/127.0.0.1:8032
19/03/20 14:26:44 INFO input.FileInputFormat: Total input paths to process
: 1
19/03/20 14:26:44 INFO mapreduce.JobSubmitter: number of splits:1
19/03/20 14:26:45 INFO mapreduce.JobSubmitter: Submitting tokens for job:
job_1553060650698_0010
19/03/20 14:26:45 INFO impl.YarnClientImpl: Submitted application
application_1553060650698_0010
19/03/20 14:26:46 INFO mapreduce.Job: The url to track the job:
http://localhost:8088/proxy/application_1553060650698_0010/
19/03/20 14:26:46 INFO mapreduce.Job: Running job: job_1553060650698_0010
19/03/20 14:27:02 INFO mapreduce.Job: Job job_1553060650698_0010 running
in uber mode : false
19/03/20 14:27:02 INFO mapreduce.Job:  map 0% reduce 0%
19/03/20 14:27:18 INFO mapreduce.Job:  map 100% reduce 0%
19/03/20 14:27:33 INFO mapreduce.Job:  map 100% reduce 100%
19/03/20 14:27:33 INFO mapreduce.Job: Job job_1553060650698_0010 completed
successfully
19/03/20 14:27:34 INFO mapreduce.Job: Counters: 49
    File System Counters
        FILE: Number of bytes read=56
        FILE: Number of bytes written=215647
        FILE: Number of read operations=0
        FILE: Number of large read operations=0
        FILE: Number of write operations=0
        HDFS: Number of bytes read=980
        HDFS: Number of bytes written=33
        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)=10590
        Total time spent by all reduces in occupied slots (ms)=11672
        Total time spent by all map tasks (ms)=10590
        Total time spent by all reduce tasks (ms)=11672
        Total vcore-milliseconds taken by all map tasks=10590
        Total vcore-milliseconds taken by all reduce tasks=11672
        Total megabyte-milliseconds taken by all map tasks=10844160
        Total megabyte-milliseconds taken by all reduce tasks=11952128
    Map-Reduce Framework
        Map input records=1
        Map output records=417
        Map output bytes=6116
        Map output materialized bytes=56
```

```

    Input split bytes=117
    Combine input records=417
    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)=2580
    Physical memory (bytes) snapshot=283627520
    Virtual memory (bytes) snapshot=4118224896
    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=863
File Output Format Counters
    Bytes Written=33
The standard deviation is: 2.6011945196269792
You have new mail in /var/spool/mail/acadgild

[acadgild@localhost ~]$ hadoop fs -ls /user/acadgild/Hadoop2
19/03/20 14:28:42 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 2019-03-20 14:27
/user/acadgild/Hadoop2/_SUCCESS
-rw-r--r--    1 acadgild supergroup        33 2019-03-20 14:27
/user/acadgild/Hadoop2/part-r-00000
You have new mail in /var/spool/mail/acadgild

[acadgild@localhost ~]$ hadoop fs -cat
/user/acadgild/Hadoop2/part-r-00000
19/03/20 14:29:01 WARN util.NativeCodeLoader: Unable to load native-hadoop
library for your platform... using builtin-java classes where applicable
count 139
length      720
square      4670

```

```
[acadgild@localhost ~]$ hadoop fs -cat /user/acadgild/Hadoop2/part-r-00000
19/03/20 14:29:01 WARN util.NativeCodeLoader: Unable to load native-hadoop libra
ry for your platform... using builtin-java classes where applicable
count  139
length 720
square 4670
[acadgild@localhost ~]$
```

So Word Standard Deviation is: 2.6011945196269792