

Assignment -3

Task 1:

Execute WordMedian , WordMean , WordStandardDeviation programs using

hadoop-mapreduce-examples-2.9.0.jar file present in your AcadGild VM.

Note: We have hadoop-mapreduce-examples-2.6.5.jar available in Acadgild VM and as acadgild Support Team suggest to use this jar because higher version of jar file was corrupting.

We are using test.txt file for showing WordMedian , WordMean , WordStandardDeviation:

Test.txt file showing in below image:

```
[acadgild@localhost TestHadoop]$ hadoop fs -cat /test.txt
18/04/01 10:45:16 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
This is a BDHS BDHS BDHS BDHS file
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost TestHadoop]$
```

Wordmean: Counts the average length of the words in the input files.

O/P: In test.txt file: Words = 8 and length = 27

```
[acadgild@localhost TestHadoop]$ hadoop jar /home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce/hadoop-mapreduce-examples-2.6.5.jar wordmean /test.txt /wordmeanoutput
18/04/01 09:37:19 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/04/01 09:37:24 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/04/01 09:37:31 INFO input.FileInputFormat: Total input paths to process : 1
18/04/01 09:37:32 INFO mapreduce.JobSubmitter: number of splits:1
18/04/01 09:37:33 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1522549133607_0002
18/04/01 09:37:34 INFO impl.YarnClientImpl: Submitted application application_1522549133607_0002
18/04/01 09:37:34 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1522549133607_0002/
18/04/01 09:37:34 INFO mapreduce.Job: Running job: job_1522549133607_0002
18/04/01 09:38:14 INFO mapreduce.Job: Job job_1522549133607_0002 running in uber mode : false
18/04/01 09:38:14 INFO mapreduce.Job: map 0% reduce 0%
18/04/01 09:38:46 INFO mapreduce.Job: map 100% reduce 0%
18/04/01 09:39:15 INFO mapreduce.Job: map 100% reduce 100%
18/04/01 09:39:16 INFO mapreduce.Job: Job job_1522549133607_0002 completed successfully
18/04/01 09:39:17 INFO mapreduce.Job: Counters: 49
  File System Counters
    FILE: Number of bytes read=39
    FILE: Number of bytes written=215395
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=130
    HDFS: Number of bytes written=18
    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)=27616
    Total time spent by all reduces in occupied slots (ms)=24604
    Total time spent by all map tasks (ms)=27616
    Total time spent by all reduce tasks (ms)=24604
    Total vcore-milliseconds taken by all map tasks=27616
    Total vcore-milliseconds taken by all reduce tasks=24604
    Total megabyte-milliseconds taken by all map tasks=28278784
    Total megabyte-milliseconds taken by all reduce tasks=25194496
  Map-Reduce Framework
    Map input records=1
    Map output records=16
    Map output bytes=232
```

o/p:

```
Total vcore-milliseconds taken by all reduce tasks=24684
Total megabyte-milliseconds taken by all map tasks=28278784
Total megabyte-milliseconds taken by all reduce tasks=25194496
Map-Reduce Framework
  Map input records=1
  Map output records=16
  Map output bytes=232
  Map output materialized bytes=39
  Input split bytes=95
  Combine input records=16
  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)=613
  CPU time spent (ms)=6040
  Physical memory (bytes) snapshot=297283584
  Virtual memory (bytes) snapshot=4122935296
  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=35
File Output Format Counters
  Bytes Written=18
The mean is: 3.375
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost TestHadoop]$ hadoop fs -cat /wordmedianoutput/part-r-00000
18/04/01 09:44:57 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
count      8
length     27
```

Wordmedian: Counts the median length of the words in the input files.

```
[acadgild@localhost TestHadoop]$ hadoop jar /home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce/hadoop-mapreduce-examples-2.6.5.jar wordmedian /
test.txt /wordmedianoutput
18/04/01 09:48:21 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/04/01 09:48:25 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/04/01 09:48:29 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your applicati
on with ToolRunner to remedy this.
18/04/01 09:48:31 INFO input.FileInputFormat: Total input paths to process : 1
18/04/01 09:48:31 INFO mapreduce.JobSubmitter: number of splits:1
18/04/01 09:48:32 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1522549133607_0003
18/04/01 09:48:33 INFO impl.YarnClientImpl: Submitted application application_1522549133607_0003
18/04/01 09:48:33 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1522549133607_0003/
18/04/01 09:48:33 INFO mapreduce.Job: Running job: job_1522549133607_0003
18/04/01 09:49:02 INFO mapreduce.Job: Job job_1522549133607_0003 running in uber mode : false
18/04/01 09:49:02 INFO mapreduce.Job:  map 0% reduce 0%
18/04/01 09:49:22 INFO mapreduce.Job:  map 100% reduce 0%
18/04/01 09:49:44 INFO mapreduce.Job:  map 100% reduce 100%
18/04/01 09:49:46 INFO mapreduce.Job: Job job_1522549133607_0003 completed successfully
18/04/01 09:49:47 INFO mapreduce.Job: Counters: 49
File System Counters
  FILE: Number of bytes read=36
  FILE: Number of bytes written=215129
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=130
  HDFS: Number of bytes written=12
  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)=16970
  Total time spent by all reduces in occupied slots (ms)=18864
  Total time spent by all map tasks (ms)=16970
  Total time spent by all reduce tasks (ms)=18864
  Total vcore-milliseconds taken by all map tasks=16970
  Total vcore-milliseconds taken by all reduce tasks=18864
  Total megabyte-milliseconds taken by all map tasks=17377280
  Total megabyte-milliseconds taken by all reduce tasks=19316736
Map-Reduce Framework
  Map input records=1
```

o/p:

```
[acadgild@localhost TestHadoop]$ hadoop fs -cat /wordmedianoutput/part-r-00000
18/04/01 09:54:13 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-j
1      1
2      1
4      6
You have new mail in /var/spool/mail/acadgild
```

WordStandarddeviation: Counts the standard deviation of the length of the words in the input files.

```
[acadgild@localhost TestHadoop]$ hadoop jar /home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce/hadoop-mapreduce-examples-2.6.5.jar wordstandarddeviation /test.txt /wordstandarddeviationoutput
18/04/01 10:28:34 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/04/01 10:28:37 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/04/01 10:28:43 INFO input.FileInputFormat: Total input paths to process : 1
18/04/01 10:28:43 INFO mapreduce.JobSubmitter: number of splits:1
18/04/01 10:28:44 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1522549133607_0004
18/04/01 10:28:45 INFO impl.YarnClientImpl: Submitted application application_1522549133607_0004
18/04/01 10:28:45 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1522549133607_0004/
18/04/01 10:28:45 INFO mapreduce.Job: Running job: job_1522549133607_0004
18/04/01 10:29:14 INFO mapreduce.Job: Job job_1522549133607_0004 running in uber mode : false
18/04/01 10:29:14 INFO mapreduce.Job: map 0% reduce 0%
18/04/01 10:29:33 INFO mapreduce.Job: map 100% reduce 0%
18/04/01 10:29:55 INFO mapreduce.Job: map 100% reduce 100%
18/04/01 10:29:56 INFO mapreduce.Job: Job job_1522549133607_0004 completed successfully
18/04/01 10:29:57 INFO mapreduce.Job: Counters: 49

File System Counters
  FILE: Number of bytes read=56
  FILE: Number of bytes written=215615
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=130
  HDFS: Number of bytes written=29
  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)=16273
  Total time spent by all reduces in occupied slots (ms)=18616
  Total time spent by all map tasks (ms)=16273
  Total time spent by all reduce tasks (ms)=18616
  Total vcore-milliseconds taken by all map tasks=16273
  Total vcore-milliseconds taken by all reduce tasks=18616
```

o/p:

```
[acadgild@localhost TestHadoop]$ hadoop fs -cat /wordstandarddeviation/part-r-000000
18/04/01 10:30:43 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
cat: /wordstandarddeviation/part-r-000000: No such file or directory
[acadgild@localhost TestHadoop]$ hadoop jar /home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce/hadoop-mapreduce-examples-2.6.5.jar wordstandarddeviation /test.txt /wordstandarddeviationoutput^C
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost TestHadoop]$ hadoop fs -cat /wordstandarddeviationoutput/part-r-000000
18/04/01 10:37:25 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
count      8
length    27
square    101
```