

Session3_YARN_assignment1

Task 1:

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

Solution: WordMedian

```
[acadgild@localhost mapreduce]$ hadoop jar hadoop-mapreduce-examples-2.7.2.jar wordmedian /hadoopdata1/hdfs/test.txt /hadoopdata1/wordmedian
```

Output: The median is: 6

Detailed output:

```
18/05/20 20:22:17 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
18/05/20 20:22:23 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
18/05/20 20:22:27 INFO input.FileInputFormat: Total input paths to process : 1
18/05/20 20:22:28 INFO mapreduce.JobSubmitter: number of splits:1
18/05/20 20:22:30 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1526825720506_0002
18/05/20 20:22:33 INFO impl.YarnClientImpl: Submitted application application_1526825720506_0002
18/05/20 20:22:35 INFO mapreduce.Job: The url to track the
job: http://localhost:8088/proxy/application\_1526825720506\_0002/
18/05/20 20:22:35 INFO mapreduce.Job: Running job: job_1526825720506_0002
18/05/20 20:23:52 INFO mapreduce.Job: Job job_1526825720506_0002 running in uber mode : false
18/05/20 20:23:52 INFO mapreduce.Job: map 0% reduce 0%
18/05/20 20:24:55 INFO mapreduce.Job: map 100% reduce 0%
18/05/20 20:25:49 INFO mapreduce.Job: map 100% reduce 100%
18/05/20 20:25:56 INFO mapreduce.Job: Job job_1526825720506_0002 completed successfully
18/05/20 20:25:59 INFO mapreduce.Job: Counters: 49
  File System Counters
    FILE: Number of bytes read=66
    FILE: Number of bytes written=235163
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=171
    HDFS: Number of bytes written=25
    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)=52337
    Total time spent by all reduces in occupied slots (ms)=52335
    Total time spent by all map tasks (ms)=52337
    Total time spent by all reduce tasks (ms)=52335
    Total vcore-milliseconds taken by all map tasks=52337
    Total vcore-milliseconds taken by all reduce tasks=52335
    Total megabyte-milliseconds taken by all map tasks=53593088
    Total megabyte-milliseconds taken by all reduce tasks=53591040
  Map-Reduce Framework
```

```

Map input records=2
Map output records=8
Map output bytes=64
Map output materialized bytes=66
Input split bytes=112
Combine input records=8
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)=1317
CPU time spent (ms)=8260
Physical memory (bytes) snapshot=229036032
Virtual memory (bytes) snapshot=638480384
Total committed heap usage (bytes)=137498624

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=59
File Output Format Counters
Bytes Written=25
The median is: 6

[acadgild@localhost mapreduce]$ hdfs dfs -ls /hadoopdata1/wordmedian
Found 2 items
-rw-r--r-- 1 acadgild supergroup      0 2018-05-20 20:25 /hadoopdata1/wordmedian/_SUCCESS
-rw-r--r-- 1 acadgild supergroup    25 2018-05-20 20:25 /hadoopdata1/wordmedian/part-r-00000
[acadgild@localhost mapreduce]$ hdfs dfs -cat /hadoopdata1/wordmedian/part-r-00000

3  2
5  1
6  1
7  2
8  1
12 1

```

Solution: WordMean

```

[acadgild@localhost mapreduce]$ hadoop jar hadoop-mapreduce-examples-2.7.2.jar wordmean
/hadoopdata1/hdfs/test.txt /hadoopdata1/wordmean

```

Output: The mean is: 6.375

Detailed output:

18/05/20 21:41:03 INFO client.RMPProxy: Connecting to ResourceManager at /0.0.0.0:8032
18/05/20 21:41:11 INFO input.FileInputFormat: Total input paths to process : 1
18/05/20 21:41:13 INFO mapreduce.JobSubmitter: number of splits:1
18/05/20 21:41:15 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1526825720506_0003
18/05/20 21:41:18 INFO impl.YarnClientImpl: Submitted application application_1526825720506_0003
18/05/20 21:41:19 INFO mapreduce.Job: The url to track the
job: http://localhost:8088/proxy/application_1526825720506_0003/
18/05/20 21:41:19 INFO mapreduce.Job: Running job: job_1526825720506_0003
18/05/20 21:43:14 INFO mapreduce.Job: Job job_1526825720506_0003 running in uber mode : false
18/05/20 21:43:14 INFO mapreduce.Job: map 0% reduce 0%
18/05/20 21:44:45 INFO mapreduce.Job: map 100% reduce 0%
18/05/20 21:46:00 INFO mapreduce.Job: map 100% reduce 67%
18/05/20 21:46:04 INFO mapreduce.Job: map 100% reduce 100%
18/05/20 21:46:10 INFO mapreduce.Job: Job job_1526825720506_0003 completed successfully
18/05/20 21:46:11 INFO mapreduce.Job: Counters: 49

File System Counters

FILE: Number of bytes read=39
FILE: Number of bytes written=235369
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=171
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)=79624
Total time spent by all reduces in occupied slots (ms)=73207
Total time spent by all map tasks (ms)=79624
Total time spent by all reduce tasks (ms)=73207
Total vcore-milliseconds taken by all map tasks=79624
Total vcore-milliseconds taken by all reduce tasks=73207
Total megabyte-milliseconds taken by all map tasks=81534976
Total megabyte-milliseconds taken by all reduce tasks=74963968

Map-Reduce Framework

Map input records=2
Map output records=16
Map output bytes=232
Map output materialized bytes=39
Input split bytes=112
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)=1970
CPU time spent (ms)=8930
Physical memory (bytes) snapshot=228909056
Virtual memory (bytes) snapshot=637632512
Total committed heap usage (bytes)=137498624

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=59
File Output Format Counters
  Bytes Written=18
The mean is: 6.375
[acadgild@localhost mapreduce]$
```

```
[acadgild@localhost mapreduce]$ hdfs dfs -ls
/hadoopdata1/wordmean
```

```
Found 2 items
-rw-r--r--  1 acadgild supergroup      0 2018-05-20 21:46 /hadoopdata1/wordmean/_SUCCESS
-rw-r--r--  1 acadgild supergroup    18 2018-05-20 21:46 /hadoopdata1/wordmean/part-r-00000
```

```
[acadgild@localhost mapreduce]$ hdfs dfs -cat
/hadoopdata1/wordmean/part-r-00000
```

```
count  8
length 51
```

Solution: WordStandardDeviation

```
[acadgild@localhost mapreduce]$ hadoop jar hadoop-mapreduce-examples-2.7.2.jar wordstandarddeviation
/hadoopdata1/hdfs/test.txt /hadoopdata1/wordstdeviation
```

Output: The standard deviation is: 2.7357585785299112

Detailed output:

```
18/05/20 22:07:19 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
18/05/20 22:07:32 INFO input.FileInputFormat: Total input paths to process : 1
18/05/20 22:07:34 INFO mapreduce.JobSubmitter: number of splits:1
18/05/20 22:07:38 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1526825720506_0004
18/05/20 22:07:44 INFO impl.YarnClientImpl: Submitted application application_1526825720506_0004
18/05/20 22:07:47 INFO mapreduce.Job: The url to track the
job: http://localhost:8088/proxy/application\_1526825720506\_0004/
18/05/20 22:07:47 INFO mapreduce.Job: Running job: job_1526825720506_0004
18/05/20 22:09:36 INFO mapreduce.Job: Job job_1526825720506_0004 running in uber mode : false
18/05/20 22:09:36 INFO mapreduce.Job:  map 0% reduce 0%
18/05/20 22:10:32 INFO mapreduce.Job:  map 100% reduce 0%
18/05/20 22:11:26 INFO mapreduce.Job:  map 100% reduce 100%
18/05/20 22:11:32 INFO mapreduce.Job: Job job_1526825720506_0004 completed successfully
18/05/20 22:11:34 INFO mapreduce.Job: Counters: 49
  File System Counters
    FILE: Number of bytes read=56
    FILE: Number of bytes written=235585
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=171
    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)=48194
Total time spent by all reduces in occupied slots (ms)=49634
Total time spent by all map tasks (ms)=48194
Total time spent by all reduce tasks (ms)=49634
Total vcore-milliseconds taken by all map tasks=48194
Total vcore-milliseconds taken by all reduce tasks=49634
Total megabyte-milliseconds taken by all map tasks=49350656
Total megabyte-milliseconds taken by all reduce tasks=50825216
Map-Reduce Framework
  Map input records=2
  Map output records=24
  Map output bytes=352
  Map output materialized bytes=56
  Input split bytes=112
  Combine input records=24
  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)=1184
  CPU time spent (ms)=8000
  Physical memory (bytes) snapshot=228827136
  Virtual memory (bytes) snapshot=636063744
  Total committed heap usage (bytes)=137498624
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=59
File Output Format Counters
  Bytes Written=29
The standard deviation is: 2.7357585785299112
[acadgild@localhost mapreduce]$

```

```

[acadgild@localhost mapreduce]$ hdfs dfs -ls
/hadoopdata1/wordstdeviation

```

```

-rw-r--r--  1 acadgild supergroup      0 2018-05-20 22:11 /hadoopdata1/wordstdeviation/_SUCCESS
-rw-r--r--  1 acadgild supergroup    29 2018-05-20 22:11 /hadoopdata1/wordstdeviation/part-r-00000

```

```

[acadgild@localhost mapreduce]$ hdfs dfs -cat
/hadoopdata1/wordstdeviation/part-r-00000

```

```

count  8
length 51
square 385

```