Assignment3

Task 1

[acadgild@localhost mapreduce]$ hdfs dfs -cat hadoop/word-count.txt
Java HotSpot(TM) Client VM warning: You have loaded library /home/acadgild/hadoop-2.7.2/lib/native/libhadoop.so.1.0.0 which might have disabled stack guard. The VM will try to fix the stack guard now. It's highly recommended that you fix the library with 'execstack -c <libfile>', or link it with '-z noexecstack'.
18/06/09 17:59:16 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
  GNU nano 2.0.9   File: word-count.txt   Modified

i love hadoop

[New File]
^G Get Help ^O WriteOut ^R Read File ^Y Prev Page ^K Cut Text ^C Cur Pos
^X Exit^J Justify   ^W Where Is ^V Next Page ^U UnCut Text^T To Spell

Wordcount

[acadgild@localhost mapreduce]$ hadoop jar hadoop-mapreduce-examples-2.7.2.jar wordcount hadoop/word-count.txt hadoop/word-countout1
Java HotSpot(TM) Client VM warning: You have loaded library /home/acadgild/hadoop-2.7.2/lib/native/libhadoop.so.1.0.0 which might have disabled stack guard. The VM will try to fix the stack guard now. It's highly recommended that you fix the library with 'execstack -c <libfile>', or link it with '-z noexecstack'.
18/06/09 18:05:14 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/06/09 18:05:16 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
18/06/09 18:05:17 INFO input.FileInputFormat: Total input paths to process : 1
18/06/09 18:05:17 INFO mapreduce.JobSubmitter: number of splits:1
18/06/09 18:05:18 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1528539845120_0004
18/06/09 18:05:18 INFO impl.YarnClientImpl: Submitted application application_1528539845120_0004
18/06/09 18:05:19 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1528539845120_0004/
18/06/09 18:05:19 INFO mapreduce.Job: Running job: job_1528539845120_0004
18/06/09 18:05:33 INFO mapreduce.Job: Job job_1528539845120_0004 running in uber mode : false
18/06/09 18:05:33 INFO mapreduce.Job: map 0% reduce 0%
18/06/09 18:05:42 INFO mapreduce.Job: map 100% reduce 0%
18/06/09 18:05:52 INFO mapreduce.Job: map 100% reduce 100%
18/06/09 18:05:52 INFO mapreduce.Job: Job job_1528539845120_0004 completed successfully
18/06/09 18:05:53 INFO mapreduce.Job: Counters: 49
        File System Counters
                FILE: Number of bytes read=698
                FILE: Number of bytes written=236741
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=2066
                HDFS: Number of bytes written=521
                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)=6776
        Total time spent by all reduces in occupied slots (ms)=7144
        Total time spent by all map tasks (ms)=6776
        Total time spent by all reduce tasks (ms)=7144
        Total vcore-milliseconds taken by all map tasks=6776
        Total vcore-milliseconds taken by all reduce tasks=7144
        Total megabyte-milliseconds taken by all map tasks=6938624
        Total megabyte-milliseconds taken by all reduce tasks=7315456
Map-Reduce Framework
        Map input records=24
        Map output records=63
        Map output bytes=2154
        Map output materialized bytes=698
        Input split bytes=122
        Combine input records=63
        Combine output records=43
        Reduce input groups=43
        Reduce shuffle bytes=698
        Reduce input records=43
        Reduce output records=43
        Spilled Records=86
        Shuffled Maps =1
        Failed Shuffles=0
        Merged Map outputs=1
        GC time elapsed (ms)=234
        CPU time spent (ms)=1790
        Physical memory (bytes) snapshot=229429248
        Virtual memory (bytes) snapshot=636829696
        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=1944
File Output Format Counters
        Bytes Written=521

Wordmedian

[acadgild@localhost mapreduce]$ hadoop jar hadoop-mapreduce-examples-2.7.2.jar wordmedian hadoop/word-count.txt hadoop/word-medianout
Java HotSpot(TM) Client VM warning: You have loaded library /home/acadgild/hadoop-2.7.2/lib/native/libhadoop.so.1.0.0 which might have disabled stack guard. The VM will try to fix the stack guard now. It's highly recommended that you fix the library with 'execstack -c <libfile>', or link it with '-z noexecstack'.
18/06/09 18:00:22 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/06/09 18:00:23 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
18/06/09 18:00:24 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not

performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
18/06/09 18:00:25 INFO input.FileInputFormat: Total input paths to process : 1
18/06/09 18:00:25 INFO mapreduce.JobSubmitter: number of splits:1
18/06/09 18:00:26 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1528539845120_0003
18/06/09 18:00:26 INFO impl.YarnClientImpl: Submitted application application_1528539845120_0003
18/06/09 18:00:26 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1528539845120_0003/
18/06/09 18:00:26 INFO mapreduce.Job: Running job: job_1528539845120_0003
18/06/09 18:00:42 INFO mapreduce.Job: Job job_1528539845120_0003 running in uber mode : false
18/06/09 18:00:42 INFO mapreduce.Job: map 0% reduce 0%
18/06/09 18:00:51 INFO mapreduce.Job: map 100% reduce 0%
18/06/09 18:01:02 INFO mapreduce.Job: map 100% reduce 100%
18/06/09 18:01:03 INFO mapreduce.Job: Job job_1528539845120_0003 completed successfully
18/06/09 18:01:04 INFO mapreduce.Job: Counters: 49
        File System Counters
                FILE: Number of bytes read=146
                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=2066
                HDFS: Number of bytes written=64
                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)=7700
                Total time spent by all reduces in occupied slots (ms)=7208
                Total time spent by all map tasks (ms)=7700
                Total time spent by all reduce tasks (ms)=7208
                Total vcore-milliseconds taken by all map tasks=7700
                Total vcore-milliseconds taken by all reduce tasks=7208
                Total megabyte-milliseconds taken by all map tasks=7884800
                Total megabyte-milliseconds taken by all reduce tasks=7380992
        Map-Reduce Framework
                Map input records=24
                Map output records=63
                Map output bytes=504
                Map output materialized bytes=146
                Input split bytes=122
                Combine input records=63
                Combine output records=14
                Reduce input groups=14
                Reduce shuffle bytes=146
                Reduce input records=14
                Reduce output records=14
                Spilled Records=28
                Shuffled Maps =1
                Failed Shuffles=0
                Merged Map outputs=1
                GC time elapsed (ms)=240
                CPU time spent (ms)=2090
                Physical memory (bytes) snapshot=229339136

Virtual memory (bytes) snapshot=635781120
                        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=1944
                File Output Format Counters
                        Bytes Written=64
The median is: 5

wordmean

[acadgild@localhost mapreduce]$ hadoop jar hadoop-mapreduce-examples-2.7.2.jar wordmean hadoop/word-count.txt hadoop/word-mean
Java HotSpot(TM) Client VM warning: You have loaded library /home/acadgild/hadoop-2.7.2/lib/native/libhadoop.so.1.0.0 which might have disabled stack guard. The VM will try to fix the stack guard now. It's highly recommended that you fix the library with 'execstack -c <libfile>', or link it with '-z noexecstack'.
18/06/09 18:06:14 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/06/09 18:06:16 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
18/06/09 18:06:18 INFO input.FileInputFormat: Total input paths to process : 1
18/06/09 18:06:18 INFO mapreduce.JobSubmitter: number of splits:1
18/06/09 18:06:18 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1528539845120_0005
18/06/09 18:06:19 INFO impl.YarnClientImpl: Submitted application application_1528539845120_0005
18/06/09 18:06:19 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1528539845120_0005/
18/06/09 18:06:19 INFO mapreduce.Job: Running job: job_1528539845120_0005
18/06/09 18:06:33 INFO mapreduce.Job: Job job_1528539845120_0005 running in uber mode : false
18/06/09 18:06:33 INFO mapreduce.Job: map 0% reduce 0%
18/06/09 18:06:42 INFO mapreduce.Job: map 100% reduce 0%
18/06/09 18:06:53 INFO mapreduce.Job: map 100% reduce 100%
18/06/09 18:06:54 INFO mapreduce.Job: Job job_1528539845120_0005 completed successfully
18/06/09 18:06:54 INFO mapreduce.Job: Counters: 49
        File System Counters
                FILE: Number of bytes read=39
                FILE: Number of bytes written=235413
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=2066
                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)=6653

Total time spent by all reduces in occupied slots (ms)=6796
                    Total time spent by all map tasks (ms)=6653
                    Total time spent by all reduce tasks (ms)=6796
                    Total vcore-milliseconds taken by all map tasks=6653
                    Total vcore-milliseconds taken by all reduce tasks=6796
                    Total megabyte-milliseconds taken by all map tasks=6812672
                    Total megabyte-milliseconds taken by all reduce tasks=6959104
          Map-Reduce Framework
                    Map input records=24
                    Map output records=126
                    Map output bytes=1827
                    Map output materialized bytes=39
                    Input split bytes=122
                    Combine input records=126
                    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)=257
                    CPU time spent (ms)=1770
                    Physical memory (bytes) snapshot=229376000
                    Virtual memory (bytes) snapshot=636829696
                    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=1944
          File Output Format Counters
                    Bytes Written=21
The mean is: 29.19047619047619

word standarddeviation

[acadgild@localhost mapreduce]$ hadoop jar hadoop-mapreduce-examples-2.7.2.jar wordstandarddeviation hadoop/word-count.txt hadoop/word-standarddeviation
Java HotSpot(TM) Client VM warning: You have loaded library /home/acadgild/hadoop-2.7.2/lib/native/libhadoop.so.1.0.0 which might have disabled stack guard. The VM will try to fix the stack guard now. It's highly recommended that you fix the library with 'execstack -c <libfile>', or link it with '-z noexecstack'.
18/06/09 18:09:12 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/06/09 18:09:13 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
18/06/09 18:09:15 INFO input.FileInputFormat: Total input paths to process : 1
18/06/09 18:09:15 INFO mapreduce.JobSubmitter: number of splits:1
18/06/09 18:09:15 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1528539845120_0006
18/06/09 18:09:16 INFO impl.YarnClientImpl: Submitted application application_1528539845120_0006

18/06/09 18:09:16 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1528539845120_0006/
18/06/09 18:09:16 INFO mapreduce.Job: Running job: job_1528539845120_0006
18/06/09 18:09:30 INFO mapreduce.Job: Job job_1528539845120_0006 running in uber mode : false
18/06/09 18:09:30 INFO mapreduce.Job: map 0% reduce 0%
18/06/09 18:09:40 INFO mapreduce.Job: map 100% reduce 0%
18/06/09 18:09:50 INFO mapreduce.Job: map 100% reduce 100%
18/06/09 18:09:50 INFO mapreduce.Job: Job job_1528539845120_0006 completed successfully
18/06/09 18:09:50 INFO mapreduce.Job: Counters: 49
        File System Counters
                FILE: Number of bytes read=56
                FILE: Number of bytes written=235637
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=2066
                HDFS: Number of bytes written=35
                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)=6948
                Total time spent by all reduces in occupied slots (ms)=7015
                Total time spent by all map tasks (ms)=6948
                Total time spent by all reduce tasks (ms)=7015
                Total vcore-milliseconds taken by all map tasks=6948
                Total vcore-milliseconds taken by all reduce tasks=7015
                Total megabyte-milliseconds taken by all map tasks=7114752
                Total megabyte-milliseconds taken by all reduce tasks=7183360
        Map-Reduce Framework
                Map input records=24
                Map output records=189
                Map output bytes=2772
                Map output materialized bytes=56
                Input split bytes=122
                Combine input records=189
                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)=241
                CPU time spent (ms)=1900
                Physical memory (bytes) snapshot=229335040
                Virtual memory (bytes) snapshot=635932672
                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=1944
        File Output Format Counters
            Bytes Written=35
```
The standard deviation is: 34.907068071674885