Mahout - Page Rank

Runtime

```
[ec2-user@ip-172-31-75-50 src]$ time hadoop jar PageRank.jar it.uniroma1.hadoop.
pagerank.PageRank --input /data/webStanford --output /data/prOutput --damping 90
 --count 8
Damping factor: 1.0
Number of iterations: 8
Input directory: /data/webStanford
Output directory: /data/prOutput
----------------------------
Running Job#1 (graph parsing) ...
20/11/06 01:15:12 INFO client.RMProxy: Connecting to ResourceManager at /172.31.
75.50:8032
20/11/06 01:15:12 WARN mapreduce.JobResourceUploader: Hadoop command-line option
 parsing not performed. Implement the Tool interface and execute your applicatio
n with ToolRunner to remedy this.
20/11/06 01:15:12 INFO input.FileInputFormat: Total input paths to process : 1
20/11/06 01:15:13 INFO mapreduce.JobSubmitter: number of splits:1
20/11/06 01:15:13 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_16
04624942236_0001
20/11/06 01:15:13 INFO impl.YarnClientImpl: Submitted application application_16
04624942236_0001
20/11/06 01:15:13 INFO mapreduce.Job: The url to track the job: http://ip-172-31
-75-50.ec2.internal:8088/proxy/application_1604624942236_0001/
```

```
                Input split bytes=118
                Combine input records=0
                Combine output records=0
                Reduce input groups=123967
                Reduce shuffle bytes=4681263
                Reduce input records=281904
                Reduce output records=281904
                Spilled Records=563808
                Shuffled Maps =1
                Failed Shuffles=0
                Merged Map outputs=1
                GC time elapsed (ms)=129
                CPU time spent (ms)=3830
                Physical memory (bytes) snapshot=439177216
                Virtual memory (bytes) snapshot=4256829440
                Total committed heap usage (bytes)=300417024
        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=22629726
        File Output Format Counters
                Bytes Written=7333875
DONE!

real    5m10.357s
user    0m6.917s
sys     0m0.434s
```

Section c)

hadoop fs -cat /data/prOutput/result/part-r-00000 | more

```
0.15000000596046448        75380
0.15000000596046448        155237
0.15000000596046448        75378
0.15000000596046448        155226
0.15000000596046448        155222
0.15000000596046448        155221
0.15000000596046448        75372
0.15000000596046448        125860
0.15000000596046448        47294
0.15000000596046448        155216
0.15000000596046448        155204
0.15000000596046448        155203
0.15000000596046448        47303
0.15000000596046448        155116
0.15000000596046448        125881
0.15000000596046448        47316
0.15000000596046448        75340
0.15000000596046448        47317
0.15000000596046448        125872
0.15000000596046448        75358
0.15000000596046448        75357
0.15000000596046448        47327
0.15000000596046448        47336
0.15000000596046448        155122
0.15000000596046448        47357
0.15000000596046448        155140
0.15000000596046448        47360
0.15000000596046448        155137
0.15000000596046448        75347
0.15000000596046448        155129
0.15000000596046448        68099
0.15000000596046448        229465
```