PageRank

Une initialisation avec un PR de dampling_factor / total_noeuds sur tous les noeuds

Une itérations du calcul du PR, il suffit d'itérer sur le job CalculPR

noeud\itération	1	10	20	50	75	100
0	0,1503	38,19	48,54	50,65	50,66	50,66
1	0,1505	48,45	60,77	63,28	63,29	63,29
10	0,1501	17,88	23,74	24,95	24,96	24,96
100	0,1501	18,94	25,23	26,53	26,53	26,53
1000	0,1500	1,584	2,041	2,136	2,136	2,136

On constate qu'au bout de 20 itérations les scores du PageRank s'approchent fortement du score final. Après 50 itérations le score ne varie presque plus : ex entre 75 et 100 pas de variations

```java
1  package cs.bigdata.ex52;
2
3  import org.apache.hadoop.conf.Configuration;
4
5  public class PageRank {
6      static int TOTAL_NOEUDS = 75879;
7      static Double DAMPLING_FACTOR = 0.85;
8      static int nIter = 100;
9
10     public static void main(String[] args) throws Exception {
11         // Connection a HDFS
12         Configuration conf = new Configuration();
13         conf.set("fs.defaultFS","hdfs://localhost:9000");
14
15         // Step 1- Init info
16         InitPR.main(conf);
17
18         // Step 2 - Calcul du PageRank
19
20         for (int i = 1; i <= nIter; i++)
21         {
22             CalculPR.main(conf, i);
23         }
24     }
25 }
```

Console

\<terminated\> PageRank [Java Application] /Library/Java/JavaVirtualMachines/jdk1.8.0_152.jdk/Contents/Home/bin/java (15 janv. 2018 à 13:04

            HDFS: Number of read operations=2015
            HDFS: Number of large read operations=0
            HDFS: Number of write operations=1204
    Map-Reduce Framework
            Map input records=75880
            Map output records=600256
            Map output bytes=17078177
            Map output materialized bytes=18285308
            Input split bytes=137
            Combine input records=0
            Combine output records=0
            Reduce input groups=75880
            Reduce shuffle bytes=0
            Reduce input records=600256
            Reduce output records=75880
            Spilled Records=1200512
            Shuffled Maps =0
            Failed Shuffles=0
            Merged Map outputs=0
            GC time elapsed (ms)=4
            Total committed heap usage (bytes)=3827302400
    File Input Format Counters
            Bytes Read=4473596
    File Output Format Counters
            Bytes Written=4472721

```
OMA/PLP/projects  master ✗
▶ hdfs dfs -cat pageRankOutput21/part-r-00000 | head -n 6
WARNING: An illegal reflective access operation has occurred
WARNING: Illegal reflective access by org.apache.hadoop.securit
security.krb5.Config.getInstance()
WARNING: Please consider reporting this to the maintainers of c
WARNING: Use --illegal-access=warn to enable warnings of furthe
WARNING: All illegal access operations will be denied in a futu
18/01/15 13:16:03 WARN util.NativeCodeLoader: Unable to load na
        19925.5048393553
0       48.53697199108769       39,40,41,42,43,44,45,46,47,48,4
6,97,98,99,100,101,102,103,38,37,36,4,104,105,106,107,108,109,1
124,125,126,127,128,129,130,131,132,133,134,135,136,137,138,139
1       60.77209580068582       24,21,20,17,16,11,10,5,4,3,1988
02,2470,2464,2256,2144,2065,2046,2040,2036,1778,1747,1718,1642,
5,728,727,725,723,704,694,689,684,663,649,645,625,592,586,576,5
7,426,425,424,423,422,421,420,419,418,417,416,415,414,413,412,4
5,374,373,372,371,370,369,368,367,366,365,364,363,362,361,360,3
3,322,321,320,319,318,317,316,315,314,313,312,311,310,300,297,2
2,91,90,87,86,85,83,82,81,78,77,75,74,71,68,67,66,65,63,61,58,5
10      23.740468998782667      48,50,51,61,65,66,69,72,73,74,7
```