

Trees of Paris

Les 3 questions ont été résolues avec un mapreduce différent.

1. Compte des arbres par type (*CompteArbre*)

Un job mapReduce similaire au WordCount

2. PlusGrandArbre

Dans le map on transforme en type_arbre -> taille_arbre. Le reduce prend le plus grand et renvoie type_arbre -> max(taille1, taille2, taille3, ...)

3. Plus vieil arbre

Tri des arbres grâce à une CustomKey dans un job Map, on obtient ainsi l'arrondissement -> année de plantation triée par ordre croissant d'année. Le plus vieil arbre est ainsi dans le 5^{ème} arrondissement et a été planté en 1601

Questions 1-2

Type	Nombre	Taille max
Acer	3	16.0
Aesculus	3	30.0
Ailanthus	1	35.0
Alnus	1	16.0
Araucaria	1	9.0
Broussonetia	1	12.0
Calocedrus	1	20.0
Catalpa	1	15.0
Cedrus	4	30.0
Celtis	1	16.0
Corylus	3	20.0
Davidia	1	12.0
Diospyros	4	14.0
Eucommia	1	12.0
Fagus	8	30.0
Fraxinus	1	30.0
Ginkgo	5	33.0
Gymnocladus	1	10.0
Juglans	1	28.0
Liriodendron	2	35.0

Maclura	1	13.0
Magnolia	1	12.0
Paulownia	1	20.0
Pinus	5	30.0
Platanus	19	45.0
Pterocarya	3	30.0
Quercus	4	31.0
Robinia	1	11.0
Sequoia	1	30.0
Sequoiadendron	5	35.0
Styphnolobium	1	10.0
Taxodium	3	35.0
Taxus	2	13.0
Tilia	1	20.0
Ulmus	1	15.0
Zelkova	4	30.0

Question 3

Les plus vieux arbres par arrondissement:

5ième : planté en 1601

16ième: un planté en 1772 un autre en 1782

12ième : un planté en 1784

...

Screens

```
OMA/PLP/dataset master X
▶ hdfs dfs -cat treeOutput2/part-r-00000
WARNING: An illegal reflective access operation has occurred
WARNING: Illegal reflective access by org.apache.hadoop.security.krb5.Config.getInstance()
WARNING: Please consider reporting this to the maintainers of
WARNING: Use --illegal-access=warn to enable warnings of further
WARNING: All illegal access operations will be denied in a future
18/01/14 01:19:03 WARN util.NativeCodeLoader: Unable to load
Acer 16.0
Aesculus 30.0
Ailanthus 35.0
Alnus 16.0
Araucaria 9.0
Broussonetia 12.0
Calocedrus 20.0
Catalpa 15.0
Cedrus 30.0
Celtis 16.0
Corylus 20.0
Davidia 12.0
Diospyros 14.0
Eucommia 12.0
Fagus 30.0
Fraxinus 30.0
```

```
OMA/PLP/dataset master X
▶ hdfs dfs -cat treeOutput1
WARNING: An illegal reflective access operation has occurred
WARNING: Illegal reflective access by org.apache.hadoop.security.krb5.Config.getInstance()
WARNING: Please consider reporting this to the maintainers of
WARNING: Use --illegal-access=warn to enable warnings of further
WARNING: All illegal access operations will be denied in a future
18/01/14 01:19:24 WARN util.NativeCodeLoader: Unable to load
Acer 3
Aesculus 3
Ailanthus 1
Alnus 1
Araucaria 1
Broussonetia 1
Calocedrus 1
Catalpa 1
Cedrus 4
Celtis 1
Corylus 3
Davidia 1
Diospyros 4
Eucommia 1
Fagus 8
Fraxinus 1
Ginkgo 5
Gymnocladus 1
```

```

▶ hdfs dfs -cat treeOutput3/part-r-00000 | head -n 3
WARNING: An illegal reflective access operation has occurred
WARNING: Illegal reflective access by org.apache.hadoop.security.authentication.util
security.krb5.Config.getInstance()
WARNING: Please consider reporting this to the maintainers of org.apache.hadoop.sec
WARNING: Use --illegal-access=warn to enable warnings of further illegal reflective
WARNING: All illegal access operations will be denied in a future release
18/01/14 01:18:41 WARN util.NativeCodeLoader: Unable to load native-hadoop library
cs.bigdata.ex53.PlusVieilArbre$CustomKey@3ea28e31      5 -> 1601
cs.bigdata.ex53.PlusVieilArbre$CustomKey@3ea28e31      16 -> 1772
cs.bigdata.ex53.PlusVieilArbre$CustomKey@3ea28e31      16 -> 1782

```

ArbresParis.java
CompteArbre.java
PlusGrandArbre.java
PlusVieilArbre.java

```

1 package cs.bigdata.ex53;
2
3 import org.apache.hadoop.conf.Configuration;
4
5 public class ArbresParis {
6     public static void main(String[] args) throws Exception {
7         // Connection a HDFS
8         Configuration conf = new Configuration();
9         conf.set("fs.defaultFS", "hdfs://localhost:9000");
10
11         // 1. Nombre d'arbre par espece
12         CompteArbre.main(conf);
13
14         // 2. Plus grand arbre de chaque type
15         PlusGrandArbre.main(conf);
16
17         // 3. Plus vieil arbre
18         PlusVieilArbre.main(conf);
19     }
20 }

```

Console

```

<terminated> ArbresParis [Java Application] /Library/Java/JavaVirtualMachines/jdk1.8.0_152.jdk/Contents/Ho
01:17:24 INFO mapred.LocalJobRunner:
01:17:24 INFO mapred.Task: Task attempt_local1778922319_0003_r_000000_0 is allowed to co
01:17:24 INFO output.FileOutputCommitter: Saved output of task 'attempt_local1778922319_
01:17:24 INFO mapred.LocalJobRunner: reduce > reduce
01:17:24 INFO mapred.Task: Task 'attempt_local1778922319_0003_r_000000_0' done.
01:17:25 INFO mapreduce.Job: Job job_local1778922319_0003 running in uber mode : false
01:17:25 INFO mapreduce.Job: map 100% reduce 100%
01:17:25 INFO mapreduce.Job: Job job_local1778922319_0003 completed successfully
01:17:25 INFO mapreduce.Job: Counters: 29
File System Counters
    FILE: Number of bytes read=8561
    FILE: Number of bytes written=117532
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=100668
    HDFS: Number of bytes written=6457
    HDFS: Number of read operations=71
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=28
Map-Reduce Framework
    Map input records=98
    Map output records=77
    Map output bytes=1439
    Map output materialized bytes=1599
    Input split bytes=126
    Combine input records=0
    Combine output records=0
    Reduce input groups=46
    Reduce shuffle bytes=0

```