

M2 Génie Logiciel

Rapport TD9 : Spark

Pour exécuter notre main :

```
spark-submit --master yarn --deploy-mode client --class bigdata.TPSpark  
--num-executors 4 TPSpark-0.0.1.jar <inputpath citiesfile> <inputpath regionsfile>  
<outputpath directory>
```

va lancer l'ensemble de nos solutions, que vous pouvez distinguer des messages de spark par leurs lignes **bleutées**.

[Ex1: partition number](#)

Nous avons eu des difficultés durant le premier TP car on ne parvenait pas à récupérer le nombre d'exécuteurs : `context.getConf().get("spark.executor.instances")` nous retournait une erreur. Nous l'avons contournés en le changeant en `context.getConf().getInt("spark.executor.instances", 1);` sans véritablement comprendre pourquoi la première version posait problème.

[Ex2 & 3 : valid cities & stats](#)

Nous obtenons **47980** villes valides, ce qui correspond à la valeur obtenue lors du TP3 avec MapReduce.

[Ex4 : histogram](#)

Nous avons reproduit le code que vous aviez écrit au tableau lors du TP et l'avons simplement décortiqué ensemble.

[Ex5 : Jointure](#)

Nous avons commencé par créer une classe pour stocker les différents renseignements d'une ville, et faciliter la compréhension du code.

Afin d'écrire nos résultats dans un fichier, nous avons décidé d'utiliser la méthode `saveAsTextFile`. Cependant, cette méthode écrivait les pointeurs de nos objets conteneurs. Nous avons alors dû nous passer de notre classe, et nous nous contentons désormais de passer une chaîne de caractère contenant toutes les informations excepté la clé.

Nous avons une question :

nous avons plusieurs fois tentés d'afficher le contenu d'une table avec la méthode `foreach` (exemple :

```
rdd3.foreach((Tuple2<Integer,StatCounter> t) -> {System.out.print(msg+Math.pow(10,  
t._1)+" - "+t._2.count());}); }
```

cependant, sans provoquer d'erreur, cela n'affichait rien dans le terminal. Est-ce parce que le parcours se produit alors localement sur les différents clusters et que cela s'affiche donc sur "leurs" terminaux ?