

Map / Reduce

Data science
Master 2 ISiDIS
2017 / 2018

Le but est de développer des algorithmes dans le framework Map/reduce. Votre code (java) pourra à la fois être développé dans un mode mono-machine, mais aussi mode cluster pour nos Raspberry pi.

Avant de commencer, jetez un coup d'oeil sur le tutoriel de yahoo :
<https://developer.yahoo.com/hadoop/tutorial/module4.html>

1 Anagrames

Écrire un programme map/reduce qui détermine les anagrames de mots lu dans un dictionnaire. Comme fichier dictionnaire vous pouvez utiliser le fichier les mots de langue anglaise dans `/usr/share/dict/words`.

2 Réseaux sociaux

Écrire un programme map/reduce qui détermine la liste des amis en communs de chaque pairs d'utilisateurs dans un réseaux. Pour cela, nous supposons avoir le graphes des amis stocké sous forme d'une matrice d'adjacence, où pour chaque utilisateur nous avons la liste de ses amis.

3 Poste

On dispose, au sein d'un fichier csv, provenant du site officiel du gouvernement, une liste de tous les bureaux de postes situés en île de France. On souhaite réaliser un programme map/reduce avec hadoop fournissant pour chacune de ces villes une liste de coordonnées GPS (celles de chacun des bureaux de poste associés). Plus précisément, on souhaite en sortie : `VILLE (x) : (LONG,LAT),(LONG,LAT) ...` où `x` est le nombre de bureaux de poste de la ville en question et `(LONG,LAT)` les coordonnées GPS.