Bérénice Faltrept
Kevin Baptista

M2 Génie Logiciel
Rapport TD3 : MapReduce

Exercice 1 : Filtrage

Nous avons vus que le fichier est structuré par ligne ainsi
Country,City,AccentCity,Region,Population,Latitude,Longitude
mais que toutes les villes n'indiquent pas forcément toutes les informations.

Le compte de nombre de lignes sur le fichier initial est de **3173959** et nous obtenons **47980** villes indiquant leur population.

Exercice 2 : Compteurs

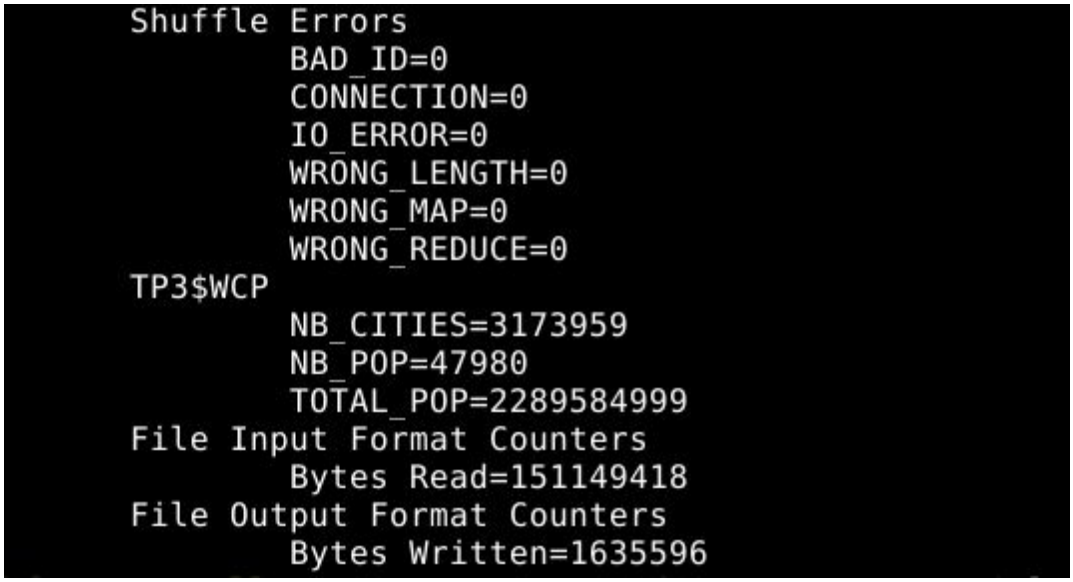
Nous créons nos compteurs dans un enum WCP :

- **NB_CITIES** qui compte le nombre de villes valide ; une ville valide est une ville qui possède un nom dans le fichier initial.
- **NB_POP** qui compte le nombre de villes avec leur population renseignée ;
- **TOTAL_POP** qui compte la population totale du fichier.

On incrémente **NB_CITIES** dans la fonction map car on doit considérer toutes les villes et pas seulement celles avec une population renseignée.
Nous avons choisi d'y placer aussi **NB_POP**, bien qu'elle aurait pu être placée au début du reducer.
On compte ensuite, dans la fonction reduce, la population totale. une fois les villes filtrées.

On utilise donc la méthode **getCounter**(Enum) dont l'argument est de la forme WCP.*type* pour accéder à un compteur spécifique.

Résultats :



Nous savons que d'autres groupes ont utilisés la méthode **getCounter**(String,String) pour parvenir à ce résultat, mais nous ne parvenons pas à comprendre où instancier les compteurs avec cette méthode, et donc utiliser les compteurs à la fois dans le Mapper et le Reducer.