

Problème 3 : Trees of Paris

Objectifs du programme Map Reduce :

- Nombre d'arbres par types (WordCount)
- Hauteur du plus haut arbre pour chaque type
- Arrondissement du plus vieil arbre dans paris

Les résultats des trois jobs sont disponibles sous le dossier **result**.

Nombre par catégorie :

Il s'agit d'un simple Wordcount, avec le type d'arbre en clé de sortie du Mapper.

Hauteur du plus haut arbre par type :

Le schéma Map Reduce est le suivant :

Mapper	<type arbre ; hauteur>
Shuffle	<type arbre ; [h1, ..., hi]>
Reducer	<type arbre ; hauteur max>

Le reducer reprend les itérables du shuffle, et possède un attribut max. Pour chaque élément, on vérifie si la valeur itérée est supérieure au max du reducer.

Arrondissement du plus vieil arbre dans Paris

Le schéma du Map Reduce est le suivant :

Mapper	<1 ; [plantation_year, borough]>
Shuffle	<1 ; [p1,b1], ..., [pi, bi]>
Reducer	<oldest_plantation_year, borough>

Il s'agit ici d'un problème de recherche de minimum, puisque nous recherchons la plus petite valeur d'année de plantation, qui correspond au plus vieil arbre.

Pour émettre des valeurs sous forme de liste, nous avons implémenté une nouvelle classe implémentant Writable.

Nous avons choisi ici d'émettre une clé unique en sortie du mapper pour que le reducer reçoive tous les éléments, et ainsi pouvoir comparer les années de plantation entre eux pour accéder à la plus petite.