

### TP 3 Programmation avec MapReduce

- ✓ Créez un dossier appelé TP3.
- ✓ Téléchargez le fichier ModeleMapReduce.tar.gz.
- ✓ Décompressez-le dans le dossier TP3.
- ✓ Lancez NetBeans. Il va vous demander dans quel workspace travailler : TP3.
- ✓ Importez le projet ModeleMapReduce dans votre Workspace
- ✓ Vous allez peut-être devoir ajouter les bibliothèques Hadoop (hadoop-common, hadoop-hdfs et hadoop-mapreduce-client-core) si vous n'avez pas fait comme demandé dans le TP2.
- ✓ Ensuite, pour créer un nouveau projet, il suffit de copier-coller le modèle.
- ✓ Vous allez écrire quelques programmes MapReduce sur le fichier des arbres remarquables de Paris. Ce fichier doit être placé sur HDFS.
- ✓ Utilisez la classe Arbre.java que vous aviez programmée dans le TP2. Cette classe regroupe toutes les fonctions de découpage et de conversion des lignes CSV en champs. Vous devrez ajouter les méthodes nécessaires pour les traitements demandés.

#### Projet 1 : Arrondissements contenant des arbres

Dupliquez le modèle de projet sous le nom **NombreArrondissements**. Écrire un programme MapReduce qui affiche la liste des arrondissements distincts contenant des arbres dans ce fichier. Forcément, c'est vingt ou moins arrondissements différents, mais combien précisément ?

Il vous suffit de placer l'arrondissement en tant que clé et de mettre une valeur quelconque ou NullWritable en tant que valeur, en sortie du mapper. Le reducer devra se contenter de sortir les clés et valeurs qu'il reçoit, il n'a même pas de boucle à faire sur les valeurs puisqu'il les ignore.

#### Projet 2 : Nombre d'arbres par genre

Dupliquez le modèle de projet sous le nom **NombreGenreArbres**. Écrire un programme MapReduce qui calcule le nombre d'arbres de chaque genre. Par exemple, il y a 3 Acer, 19 Platanus, etc. Comment allez-vous définir les clés et les valeurs transmises entre le mapper et le reducer ?

Le mapper doit extraire le genre d'arbre. Le reducer récupère les paires (clé, valeur) ayant la même clé, donc il faut que cette clé soit le genre d'arbre ; la valeur étant le nombre d'arbre de ce genre. Comme précédemment, il faut ignorer la première ligne, les titres des colonnes.

#### Projet 3 : Hauteur maximale par genre d'arbre

Dupliquez le modèle de projet sous le nom **MaxHauteurArbres**. Écrire un programme MapReduce qui calcule la hauteur du plus grand arbre de chaque genre. Par exemple, le plus grand Acer fait 16m, le plus grand Platanus fait 45m, etc.

#### Projet 4 : Arrondissement contenant le plus vieil arbre

Dupliquez le modèle de projet sous le nom **ArrondissementVieilArbre**. Écrire un programme MapReduce qui affiche l'arrondissement où se trouve le plus vieil arbre.