Atelier : Calcul du salaire Min et Max

L'objectif :

- Se familiariser avec le traitement distribué
- Utilisation du langage python pour la création des fonctions Map et Reduce
- Analyse des résultats.

Emplacement du fichier : /formation/ateliers/mapreduce/

Réalisation : Vous allez copier, supprimer plusieurs fichiers de données dans HDFS

Chapitre correspondant : HadoopDistributed File System (MapReduce)

Nous allons utiliser un fichier input sous la forme suivant:

ID ; Age ; Sexe ; Ville ; Salaire

Un exemple du contenu du fichier input :

1;22;homme;Lille;22000

2;22;homme;Dreux;14000

3;22;homme;Stratsbourg;22000

4;22;homme;Rennes;21500

5;22;homme;Nice;22000

6;22;homme;Nantes;24000

7;22;homme;Lyon;26000

8;22;homme;Chartres;20000

1- Transférer le fichier data.csv dans HDFS
2- Écrire les deux fonctions map et reduce qui ont pour but de calculer le salaire min et max par tranche par Age et le nombre d'échantillon utilisé.
    a. Tester les deux fonctions en local
    b. Une fois le test en local est ok lancer le traitement sur hadoop
        i. Avant de lancer assurez-vous que le répertoire out n'existe pas et si c'est le cas supprimez le ==> hadoop fs -rm -r out
    c. Lire le résultat sur Hadoop
3- Écrire les deux fonctions map et reduce qui ont pour but de calculer le salaire min et max par Ville et le nombre d'échantillon utilisé.
    a. Tester les deux fonctions en local
    b. Une fois le test en local est ok lancer le traitement sur hadoop
        i. Avant de lancer assurez-vous que le répertoire out n'existe pas et si c'est le cas supprimez le ==> hadoop fs -rm -r out
    c. Lire le résultat sur Hadoop

4- Écrire les deux fonctions map et reduce qui ont pour but de calculer le salaire moyen par ville et par tranche d'age.
    a. Tester les deux fonctions en local
    b. Une fois le test en local est ok lancer le traitement sur hadoop
        i. Avant de lancer assurez-vous que le répertoire out n'existe pas et si c'est le cas supprimez le ==> hadoop fs -rm -r out
    c. Lire le résultat sur Hadoop