

Atelier : Total des ventes par magasin

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:

date | temps | magasin | produit | cout | paiement

Un exemple du contenu du fichier input :

```
15/10/2019|16:40|paris|ballon |1000|cb
15/10/2019|11:49|nice|t-shirt|3000|cash
15/10/2019|13:07|bordeaux|ballon |1500|cheque
15/10/2019|10:02|paris|soulier|1900|cash
15/10/2019|16:40|dreux|xbox|350|cb
15/10/2019|11:49|lyon|iphone|999|cash
15/10/2019|13:07|lille|montre|150|cheque
15/10/2019|10:02|nante|jus orange|2|cash
15/10/2019|16:40|rennes|eau|4|cb
15/10/2019|11:49|nancy|ballon |1000|cb
```

- 1- Transférer le fichier data.csv dans HDFS
- 2- Écrire les deux fonctions map et reduce qui ont pour but de calculer le total des ventes par magasin
 - 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