

# Application du paradigme MapReduce aux données ouvertes

## Cas : Accessibilité des personnes à mobilité réduite aux musées

Billel Arres, Nadia Kabachi, Fadila Bentayeb, Omar Boussaid

Université de Lyon (Laboratoire ERIC)  
5 avenue Pierre Mandès-France, 69676 Bron, France  
{billel.arres | nadia.kabachi | fadila.bentayeb | omar.boussaid} @univ-lyon2.fr

**Résumé.** Le modèle *MapReduce* est aujourd'hui l'un des modèles de programmation parallèle les plus utilisés. Définissant une architecture Maître-Esclave, il permet le traitement parallèle de grandes masses de données. Dans ce papier, nous proposons un algorithme basé sur *MapReduce* qui permet, à partir des données publiques du Ministère Français de la Communication et de la Culture, de définir un classement des galeries et musées nationaux selon leur degré d'accessibilité aux personnes handicapées. Tout en profitant de la puissance et de la flexibilité du paradigme *MapReduce*, les décideurs pourront mettre en place des stratégies efficaces à moindre coût et avoir ainsi une vision plus précise sur les établissements culturels et leurs limites relatives à cette catégorie de personnes. L'algorithme que nous proposons peut être exploité et appliqué à d'autres cas d'études avec des jeux de données plus volumineux.

## 1 Introduction

Le monde des musées aujourd'hui connaît un engouement sans précédent qui a vu plus de 31 millions de visiteurs se précipiter en 2012 dans les musées nationaux français (NuitDesMusées, 2013). Par ailleurs, et depuis plus d'une dizaine d'années, les ministères et les services publics des différents pays accordent de plus en plus d'importance à l'ouverture et à la réutilisation de leurs données collectées ou produites au niveau de leurs différents établissements. Or, avec le temps ces données s'accumulent et deviennent très difficiles à stocker, à traiter et à analyser. D'où la nécessité de s'orienter vers de nouvelles solutions et paradigmes de programmation afin de faire face à ces difficultés. Nous proposons dans ce travail un algorithme basé sur le paradigme *MapReduce* (Dean et Ghemawat, 2008) qui permet, à partir des données ouvertes du Ministère français de la communication et de la culture, de déterminer le degré d'accessibilité des personnes handicapées (tout type d'handicape) aux musées nationaux français. L'algorithme définit, à partir de ces données brutes et hétérogènes collectées, une note pour chaque musée selon son degré d'accessibilité, et retourne un classement final des musées par commune et par région de France. Notre algorithme de classement présente plusieurs avantages notamment en terme de rapidité de traitement de grandes quantités de données puisqu'il