

# Nous recherchons un stagiaire

## Infrastructure Data Lake de zéro en héros!

**Durée:** 6 mois

### **Responsabilités:**

- Comprendre les fondamentaux de monitoring et de logging
- Définir l'architecture du lac de données à partir de zéro
- Mettre en œuvre l'infrastructure (Spark, Kafka...)
- Alimenter le lac (collecter et charger les données brutes)
- Mettre en place des outils de reporting
- Automatiser les flux de données et les rapports

### **Compétences requises:**

- Connaissances en traitement des données massives (Big Data) : Hadoop, Spark, AWS S3 (...), SQL/NoSQL.
- Construction et implementation d'une infrastructure (Spark, Kafka...)
- Connaissances de Linux
- Compétences en virtualisation

