LES TACHES & OUTILS:

* Réaliser les ordonnancements des traitements
  + Etre garant de la mise en place, du suivi et de l’exploitation des outils déployés

 la manipulation de données avec des scripts (SQL) à et du code (Python)  
- la construction et l'exploitation de pipeline de données sur des infrastructures cloud  
- l'utilisation de frameworks de calcul distribué

* Assurer l’alimentation du dataware

Au sein de l’équipe de développement, implémenter les nouvelles fonctionnalités, maintenir l’application (API, analyse documentaire, scoring multi niveaux)

Participer aux choix d’implémentations et technologiques éventuels.

Développer et construire de nouveau pipeline de données pour récupérer la Data de n'importe où et être en mesure de l'analyser

Faire évoluer et scaler la plateforme existante au sein d'un environnement Big Data avec Python

# 1

* Développer des pipelines de données, de l'ingestion jusqu'à la restitution sous format consommable par les data scientists.

**Qui êtes-vous ?**

Issu(e) d’une formation supérieure**Bac +5** ou plus (école d’ingénieurs, doctorat, master, master spécialisé), nous recherchons des **collaborateurs avec différents niveaux de séniorité** tels que des Consultants, Senior Consultants, Lead Consultants, Project Managers, …

**Votre quotidien :**

En tant que Data Engineer, vous serez amené(e) à :

* Accompagner des équipes métiers dans leurs travaux d’identification et d’expression des besoins sur la data,
* Concevoir des solutions nécessitant l’utilisation de technologies Cloud & Data pour répondre à des cas d’usages métiers,
* Développer et tester des flux d’ingestion de données de Data Lake et de traitement des données,
* Déployer, optimiser et industrialiser au quotidien des pipelines de données dans le Cloud,
* Travailler en collaboration pour soutenir le développement de nouvelles fonctionnalités.

Vos compétences techniques

* Vous possédez des connaissances liées au Cloud
* Vous possédez des compétences avancées autour des langages SQL, Python ou Scala,
* Vous êtes familier du Framework Spark,
* Vous avez une bonne maîtrise de l’optimisation des déploiements (Docker, Serverless...),
* Vous détenez des connaissances autour des processus d’industrialisation (Terraform) et des outils d’orchestration

Dans l’idéal vous avez déjà réalisé un déploiement au sein du Cloud Public (AWS, GCP, AZURE).

**Qualifications**

Vous êtes Data Engineer et avez réalisé des projets de déploiement au sein du Cloud Public. Vous aimez partager activement autour de la veille technologique, vous souhaitez transmettre votre savoir lors de sessions de formations techniques internes ou externes, mais également continuer à progresser en partageant avec des experts data internationaux autour d’ateliers techniques et en intervenant autour de sujets d’expertises, alors rejoignez-nous !

OUTILS

Développement :**Python, R, Java, Scala**Framework : **Hadoop, Spark**Outils Big data : **Yarn, Pig, Hive, Kafka, Splunk**Bases de données : **MongoDB, HBase, Cassandra**ETL : **Talend, ODI, Stambia**Plateforme : **Hortonworks, Cloudera, Map Reduce**, **AWS, GCP, Azure**