

## Poste : Data Engineer

- **Type de contrat** : Contrat
  - **Date de début** : Dès que possible
  - **Durée du mandat** : 12 mois
  - **Date limite de dépôt** : 6 novembre à 14h
- 

## À propos de l'entreprise :

**RetailNova** est un leader du secteur du commerce de détail, avec un réseau de distribution omnicanal présent à travers tout le territoire. L'entreprise combine innovation technologique et excellence opérationnelle pour optimiser ses processus, anticiper les besoins des clients et maximiser la performance. Dans le cadre de sa stratégie de transformation numérique, RetailNova investit massivement dans la modernisation de ses infrastructures de données pour accélérer la prise de décision fondée sur des analyses avancées.

---

## Contexte du mandat :

RetailNova met en œuvre une nouvelle plateforme d'entreprise basée sur des technologies cloud afin de centraliser, sécuriser et valoriser ses données opérationnelles, transactionnelles et clients. Ce projet stratégique vise à moderniser l'écosystème analytique, à automatiser les flux de données, à améliorer la gouvernance, et à renforcer la qualité et la traçabilité des données à l'échelle de l'entreprise.

Le Data Engineer sera responsable de la conception, du développement, de l'optimisation et du déploiement des pipelines de données nécessaires à l'alimentation du data lake d'entreprise. Il interviendra également dans la mise en place de nouvelles architectures de traitement en temps réel, en assurant la robustesse, la scalabilité et la performance des solutions déployées.

---

## Responsabilités détaillées :

- Concevoir et développer des pipelines d'ingestion de données robustes et optimisés à l'aide d'**Azure Data Factory** pour intégrer des données provenant de sources multiples (ERP, CRM, POS, e-commerce, etc.).

- Développer des transformations complexes en utilisant **Apache Spark** (Scala/Python) dans **Azure Databricks**, avec une forte orientation performance et optimisation des coûts.
  - Implémenter des flux de données **streaming** et **batch** via **Databricks Streaming** et **Apache Kafka**, en assurant une faible latence et une haute disponibilité.
  - Participer à la définition et à la mise en œuvre de **modèles de données** adaptés aux besoins analytiques (data marts, data lakehouse).
  - Mettre en œuvre des processus de **validation de la qualité des données** (data quality checks, data profiling, monitoring automatisé).
  - Automatiser les déploiements et les tests des pipelines via **Azure DevOps** (CI/CD, versioning, rollback).
  - Documenter les solutions développées selon les normes internes, et assurer la **traçabilité** et la **réplicabilité** des traitements.
  - Travailler en collaboration avec les équipes **Data Science, BI et IT Infrastructure** pour garantir l'alignement des livrables avec les objectifs stratégiques.
  - Participer à l'élaboration de normes et de **standards d'ingénierie des données** à l'échelle de l'organisation.
- 

### Compétences techniques requises :

- **5+ ans** d'expérience en développement d'**applications de traitement de données distribuées** avec **Apache Spark** (obligatoire).
- **Solide maîtrise de Python et Scala** pour le développement de traitements de données volumineuses.
- Expérience significative avec **Azure Data Factory, Azure Data Lake Storage (Gen2), et Azure Databricks**.
- Connaissance approfondie de **Kafka** (installation, configuration, gestion des topics, sécurité, monitoring).
- Maîtrise des **principes d'architecture data** : partitionnement, indexation, optimisation des formats (Parquet, Delta Lake), gestion des métadonnées.
- Expérience avec les outils de **gestion des flux CI/CD** dans **Azure DevOps** (pipelines, artefacts, gestion de configuration).

- Connaissance des **principes de gouvernance des données**, de **catalogage** (ex. Azure Purview), et des bonnes pratiques de **sécurité** (RBAC, encryption, masking).
  - Capacité à concevoir des **modèles de données analytiques** (Kimball, Data Vault) et à travailler avec des entrepôts de données (ex. **Synapse Analytics**).
- 

#### **Atouts :**

- Expérience dans des environnements **retail** ou **grande distribution** (gestion des données transactionnelles, gestion de la supply chain, etc.).
  - Expérience avec des **outils de monitoring** des pipelines (ex : Azure Monitor, Log Analytics, Datadog).
  - Connaissances de base en **Data Science/Machine Learning** pour interagir avec les équipes concernées.
- 

#### **Compétences interpersonnelles :**

- Forte capacité à résoudre des problèmes complexes de manière autonome.
  - Rigueur et attention aux détails, particulièrement dans les aspects liés à la qualité et la fiabilité des données.
  - Compétences en **communication technique**, capacité à documenter et présenter des solutions techniques à divers niveaux d'interlocuteurs.
  - Esprit d'équipe, ouverture au partage de connaissances et participation active à l'amélioration continue.
- 

#### **Langues :**

- **Français et anglais** : maîtrise professionnelle complète, tant à l'oral qu'à l'écrit.