

## SENIOR DATA ENGINEER - 6 ANS D'EXPÉRIENCE

### Résumé

Freelance Data Engineer expérimenté avec 6 ans d'expérience principalement dans le secteur bancaire. Expert en technologies Big Data, cloud computing et pratiques DevOps. Capable de concevoir et d'implémenter des pipelines de données complexes. Collaborateur efficace au sein d'équipes pluridisciplinaires. **Disponible pour des missions freelance ASAP.**

**Compétences clés :** Python Ecosystem (Pandas, Poetry, FastAPI, Pytest..), Spark/Databricks, Docker, Git, SQL, AWS

### Expériences professionnelles

#### Owkin

Paris - France (Full-Remote)

SENIOR DATA ENGINEER - FREELANCE (FULL REMOTE)

Août 2023 - Aujourd'hui

- Faisant partie de l'équipe "DataOps", j'ai participé aux efforts d'Owkin liés à l'extraction, à la curation et à l'alignement des données médicales brutes, et à la fourniture d'un ensemble de données adapté à l'IA, pour le compte des data scientists.
- Établir des pratiques harmonisées pour la gestion des données, spécifiquement pour gérer la qualité des données et leur livraison pour des projets internes et externes.
- Participer à l'intégration des données médicales tout au long de leur cycle de vie (gouvernance, curation, traitement).
- Développer un schéma de métadonnées pour soutenir les différents projets.
- Préparer des rapports détaillés résumant l'activité de curation des données.
- Extraction, nettoyage, curation, imputation, pseudonymisation, harmonisation, jointure de différents ensembles de données.
- Mettre en place et maintenir des pipelines CI/CD pour automatiser les tests et le déploiement, garantissant ainsi la qualité de la livraison.
- Fournir un mentorat et un coaching personnalisés aux membres juniors de l'équipe pour les aider à développer leurs compétences techniques et la prise en main des outils.
- Superviser les projets menés par les membres juniors de l'équipe, en assurant un suivi régulier et en fournissant des retours constructifs.

**TECHNOLOGIES :** Python, Poetry, Pydantic, Pandas, GH Actions, DevOps.

#### Natixis Corporate & Investment Banking

Charenton-le-Pont - France

SENIOR DATA ENGINEER - FREELANCE (2J/3J)

Février 2022 - Juin 2023

- Conception, mise en place et maintenance des pipelines de données pour la collecte, le traitement et le stockage des données de financement.
- Développement et optimisation des processus ETL pour garantir l'intégrité et la qualité des données.
- Collaboration étroite avec les équipes métier pour comprendre leurs besoins en matière de données et proposer des solutions techniques adaptées pour améliorer les processus de reporting et d'analyse.
- Encadrement et mentorat des membres plus juniors de l'équipe et partage des bonnes pratiques.

**TECHNOLOGIES :** Python, Spark, Hadoop, XLSRelease, XLSDeploy, Jenkins, Bitbucket, Control-M, Indexima.

#### La Société Générale

Fontenay-sous-Bois - France

SENIOR DATA ENGINEER - CONSULTANT DAVIDSON CONSULTING (2J/3J)

Mars 2019 - Février 2022

- Au sein du centre de compétences Big data (CCBD) puis ITIM/CFT (Fast-IT), et dans le cadre d'un projet de lutte contre la fraude interne, j'avais pour tâche d'industrialiser des traitements de détection de fraude dans un environnement big data.
- Industrialisation de traitements de détection de la fraude interne préparés par des data scientists.
- Développement de briques de traitement et d'enrichissement de données en PySpark.
- Mise en place de collecte de données depuis des datacenters de la banque de détail vers le cluster Hadoop.
- Accompagnement d'une équipe de data scientist dans la prise en main de la plateforme big data.
- Assurer la conformité RGPD au niveau de l'applicatif.
- Assurer le RUN de plusieurs application en production et accompagner les intégrateurs dans la gestion des incidents et des reprises des traitements.

**TECHNOLOGIES :** Hadoop (Hortonworks), Apache Spark, Python, Scripting Bash, Oozie, Hive, Talend, Control-M.

**Danone Research**  
DATA ENGINEER - CONSULTANT DAVIDSON CONSULTING (1J/4J)

Palaiseau - France  
Septembre 2018 - Février 2019

- Au sein de l'équipe "Biometry & Data Science" et dans le cadre du programme SMART DATA de Danone, j'avais pour rôle de mettre en place des flux d'intégration de données.
- Conception, modélisation et mise en place de processus d'intégration de données.
- Mise en place d'une analyse d'impact relative à la protection des données (DPIA) afin d'être conforme à la RGPD.
- Développement d'un portail Self-Service permettant aux équipes non techniques d'accéder aux données provenant de multiples API..
- Coordination avec des équipes à Utrecht et à Singapour en faisant le lien entre les deux entités pour assurer le bon déroulement du projet et vérifier l'adéquation de la solution avec le besoin.

**TECHNOLOGIES :** Hadoop (Cloudera), Apache Spark, Python, Pandas, Alteryx, Impala.

**Davidson Consulting**  
CONSULTANT BIG DATA - CDI

Boulogne Billancourt - France  
Septembre 2018 - Février 2022

- En tant que consultant big data, j'avais pour mission d'accompagner les différents clients de Davidson dans la mise en place de leurs projets data dans des environnements hautement distribués.
- Participation dans la préparation et la formulation des volets techniques des réponses aux appels d'offres (RAO).
- Accompagnement des ingénieurs d'affaires de Davidson dans l'identification et le recrutement de nouveaux talents et dans le développement de leurs portefeuilles clients.
- Animer des workshops visant à vulgariser plusieurs notions techniques auprès des commerciaux.

**Université Paris-Saclay**  
ENSEIGNANT VACATAIRE - CDD

Évry - France  
Septembre 2019 - Juin 2023

- Responsable du cours OLAP & bases de données multidimensionnelles du master 2 IMSD.
- Mentorat des étudiants et accompagnement de certains d'entre eux en tant que tuteur académique.

**Jems DataFactory / Laboratoire Images, Signaux et Systèmes Intelligents (LISSI)**  
DATA INGÉNIEUR R&D - ALTERNANCE

Neuilly-sur-Seine - France  
Décembre 2017 - Août 2018

- Mise en place de plusieurs uses cases autour des technologies Big Data au sein du JEMS DataLab & Le laboratoire images, signaux et systèmes intelligents (LISSI) de l'UPEC.

**TECHNOLOGIES :** Hadoop (Hortonworks), ELK Stack, NiFi, Spark, Talend.

Cursus académique

**Master 2 IMSD - Innovation, Marché et Sciences des données**  
UNIVERSITÉ PARIS SACLAY : FORMATION EN ALTERNANCE

Évry - France  
2018

- Master 2 en alternance en data sciences & data engineering.
- **DOMAINES D'EXPERTISE :** Technologies Big Data, ElasticSearch, Hadoop, Spark, NoSQL, Microsoft Azure, Machine Learning, Data Visualisation.

**Diplôme d'ingénieur en informatique (Systèmes d'Information)**  
ENICARTHAGE, ÉCOLE NATIONALE D'INGÉNIEURS DE CARTHAGE

Tunis - Tunisie  
2017

- École nationale d'ingénieurs, admission sur concours national (post-classes préparatoires).
- **DOMAINES D'EXPERTISE :** Ingénierie des systèmes d'information, développement logiciel (Java/J2EE), architecture logicielle.

Compétences

**TECHNOLOGIES :**  
**Data Engineering :** Spark, Airflow, Elastic Stack, Écosystème Hadoop, Kafka.  
**Tooling / Devops :** Git, Make, Github Actions, Jenkins, Ansible.

**Programmation :** Python, Bash scripting.  
**Bases de données :** PostgreSQL, MySQL, Hive, Indexima.

**LANGUES :**  
**Français :** Bilingue  
**Anglais :** Bilingue - TOEIC (985 / 990) ≈ 99%.

Certifications

Juin 2024  
Juin 2024

**Astronomer Certification for Apache Airflow Fundamentals**  
**Test certif??**

Astro. Academy  
Astro. Academy

# Informations Supplémentaires

---

- **CONTRAT SOUHAITÉ** : Recherche exclusivement des opportunités en Freelance à temps plein. Non intéressé par des contrats en CDI.
- **RYTHME DE TÉLÉTRAVAIL** : 2.5 jours de teletravail par semaine minimum.
- **MOBILITÉ GÉOGRAPHIQUE** : Région Île-de-France.