



Mamadou DIAKHABY

Data Engineer

Python | SQL | Airflow | dbt | Docker | FastAPI | PySpark | Kafka | GitLab CI/CD | AWS (S3)

PROFIL

“ Du laboratoire scientifique aux pipelines de données : rigueur, curiosité et automatisation. Fort de 4 ans en laboratoire (MOLP/MET), j'applique aujourd'hui la même précision expérimentale au Data Engineering : création de pipelines ETL/ELT robustes, scalables et automatisés.

Mon objectif : faire de vos données un levier de décision fiable et stratégique. ”

📞 06 03 70 08 96

✉️ diakhaby14@gmail.com

📍 Paris, Ile-de-France

in [@mamadou-diakhaby-879856156/](https://www.linkedin.com/in/mamadou-diakhaby-879856156/)

🌐 diakhaby-data.github.io

Github [Github.com/Diakhaby-data](https://github.com/Diakhaby-data)

COMPÉTENCES

- ETL/ELT, Data Warehouse, Data Pipeline, Data Modeling
- Langages & Systèmes : Python, SQL, Linux, Bash, Git, GitLab
- Bases de données : MySQL, Snowflake, MongoDB
- Big data & Streaming : PySpark, Apache Kafka, Spark
- Streaming ETL & Orchestration : Apache Airflow, dbt, automatisation de pipelines de données.
- Cloud & DevOps : Docker, Kubernetes, GitLab CI/CD, AWS (S3)
- API & Industrialisation : FastAPI, tests unitaires, conteneurisation et déploiement.
- Rigueur méthodologique • Résolution de problèmes • Autonomie

FORMATIONS

DATA ENGINEER

DataScientest - 400h, Paris, France - Juillet 2025 / Octobre 2025

- Formation de 400h centrée sur la pratique et les projets concrets.
- Construction de pipelines ETL/ELT complets, intégration et orchestration de données sur le cloud.
- Application des bonnes pratiques DevOps, tests et déploiements continus.
- Approche collaborative (Git, CI/CD) et focus sur la qualité, la performance et la fiabilité des données.

MASTER - RSE

Université de Versailles, Paris, France - 2018 / 2019

- Une formation axée sur la RSE et le développement durable.

MASTER - GÉOLOGIE

Université de Rennes 1, France - 2016 / 2018

- Prospection des Ressources Minérales et Pétrolières - RMP.

LANGUES

Français (courant), Anglais (B1), Diakhanké (langue maternelle)

INTÉRÊTS

DATA & IA

Veille technique (NLP, pipelines ETL), participation à meetups Data/ML, lecture d'articles.

LECTURE & RÉFLEXION

Essais sur technologie et Science, éthique de l'IA et innovation.

SPORT

Fitness régulier pour le maintien de la forme, résilience mentale et performance.

PROJET

Satisfaction Client - ShowroomPrivé - 230h

- Ingestion multi sources : scraping 20 259 avis (BeautifulSoup) + stockage MongoDB --> validation 98,7% d'intégrité des données.
- Pipeline ETL automatisé : orchestration Airflow (DAG de 7 tâches), transformations MySQL + Random Forest (89% accuracy) --> réduction 75% du temps manuel.
- API & observabilité production : FastAPI (latence <200 ms), dashboard Streamlit interactif, monitoring Prometheus/Grafana avec alerting automatisé.
- Industrialisation & DevOps : microservices Docker Compose (4 conteneurs), CI/CD GitHub Actions, tests unitaires (87% couverture).
- Impact business : 28k+ avis traités automatiquement, 340 avis critiques détectés, 12h/semaine libérées, architecture scalable 100k+ avis.
-
- Environnement : Python, SQL (MySQL), MongoDB, Airflow, FastAPI, Docker, Streamlit, Prometheus/Grafana, Git/GitHub.

EXPÉRIENCES

Analyste en laboratoire - Matériaux & Air (MOLP / MET) et Formateur

ITGA Analytics, Paris, France - Depuis 2022

- Analyse de matériaux de construction et filtres d'air pour détection des fibres d'amiante via Microscopie Optique en Lumière Polarisée (MOLP) et Microscopie Électronique à Transmission (MET).
- Production & qualité : analyse de 13 000 échantillons/an (1 100/mois avec respect strict protocoles ISO, taux de conformité > 99%).
- Gestion de données complexes : traçabilité exhaustive de milliers d'échantillons, détection de patterns anormaux, reporting précis sous contraintes réglementaires.
- Formation & transmission : montée en compétence des nouveaux analystes sur protocoles techniques et administratifs.
-
- Compétences transférables vers Data Engineering : rigueur méthodologique (protocoles stricts = code quality), analyse de données complexes (échantillons = datasets), gestion qualité (contrôles = tests unitaires), documentation exhaustive.