



IMAD BOUZKRAOUI

Data Engineer

+33 6 18 02 50 36 bouzkraoui.imad.2@gmail.com Île-de-France

[LinkedIn](#) [GitHub](#)

PROFIL

Data Engineer spécialisé dans la conception de pipelines distribués (ETL/ELT) et l'orchestration de workflows batch & streaming (Airflow, Prefect, Kafka, Spark). Expérience sur projets SaaS et bancaires sur AWS (ingestion, transformation, valorisation de données multi-sources) et maîtrise de la Modern Data Stack (dbt, Snowflake) via un pipeline analytics end-to-end. Attaché à la qualité, au monitoring et à la gouvernance des données, à l'aise jusqu'au dashboard décisionnel (Power BI). Recherche un CDI orienté pipelines robustes et fiables.

FORMATION

Master spécialisé en Ingénierie des Données – Institut F2I / DSP, Paris

2025 – Présent

Diplôme Ingénieur d'État, Data Engineering – ENSA AI Hoceima

2020 – 2025

EXPÉRIENCES PROFESSIONNELLES

Data Engineer – Alternance

VirtuoSys, Paris – SaaS LegalTech | Déc. 2025 – Présent

- Conception de pipelines Airflow sur AWS (S3, EC2) consolidant logs, données clients et métriques d'usage de **5 applications SaaS hétérogènes**.
- Intégration Power BI ↔ API Sentry** : dashboard self-service temps réel (3 pages, 5 projets, ~12 KPIs, RGPD) remplaçant la remontée manuelle des bugs — détection ramenée de plusieurs heures à quelques minutes.
- Scraping et ingestion de contacts publics (LegalTech) et conception d'un **CRM interne** pour structurer et exploiter ces données ; contribution à un pipeline d'intégration de chatbot.
- Industrialisation des déploiements via **GitLab CI/CD** ; pipeline d'automatisation de contenu SEO assisté par IA ; documentation et gouvernance des données.

Stack : Airflow, AWS (S3, EC2), Python, SQL, Power BI (DAX), API Sentry, GitLab CI/CD, Docker, Django, React.

Stage PFE – Data Engineer / Data Quality

Banque Centrale Populaire, Casablanca – Data Watchdog (entité Data Governance) | Fév. – Août 2025

- Conception de « **Data Watchdog** » : passage de contrôles par règles codées à une détection enrichie par **modèles ML de classification d'anomalies** (Scikit-learn).
- Orchestration **batch & streaming** (Prefect, Kafka), alertes temps réel et **boucle de validation humaine** (acceptation/rejet des corrections par les métiers) pour l'amélioration continue.
- Gouvernance & sécurité** : contrôle des accès (Apache Ranger) et lineage/catalogage des données (Apache Atlas) ; dashboards Power BI de reporting qualité (complétude, fraîcheur, anomalies).

Stack : Python, SQL, Prefect, Kafka, PostgreSQL, Scikit-learn, Apache Ranger, Apache Atlas, Power BI, Docker.

Stage PFA – Data Engineer / Deep Learning

Nextronic – Vidéosurveillance intelligente | Juin – Août 2024

- Pipeline vidéo temps réel (flux caméra live ~30 fps) classifiant **5 types d'incidents** (incendie, intrusion...) ; alerte déclenchée sur détection confirmée (persistance 3 s + scoring).
- Constitution du jeu de données (Kaggle + scraping par type d'incident) et **fine-tuning** d'un modèle pré-entraîné (Hugging Face).

Stack : Python, OpenCV, TensorFlow, Hugging Face, NumPy.

COMPÉTENCES TECHNIQUES

Langages	Python (Pandas, PySpark, NumPy), SQL / PL-SQL
Data Engineering	Airflow, Prefect, Kafka, Spark (PySpark), dbt, ETL/ELT, batch & streaming
Stockage & Architecture	Snowflake, PostgreSQL, MongoDB, MySQL, Data Lake (S3), Lakehouse, Medallion
Cloud & DevOps	AWS (S3, EC2), Docker, Git, GitLab CI/CD, Linux
Gouvernance & Qualité	Apache Atlas, Apache Ranger, qualité des données, monitoring & alerting, Sentry, RGPD
BI & Visualisation	Power BI (Desktop, Service, DAX), reporting automatisé
ML & Dev (compl.)	Scikit-learn, Hugging Face, TensorFlow, OpenCV, LLM / RAG ; Django, React/TypeScript

PROJETS

Pipeline Analytics Engineering end-to-end – Airbnb

Projet académique

dbt (Core/CLI), Snowflake, AWS (S3, IAM), Jinja, Git, SQL

- Ingestion & Data Warehouse** : données brutes via AWS S3 (storage integrations, external stages sécurisés) vers Snowflake, structurées en **architecture Medallion** (raw → silver → gold).
- Transformation dbt** : modèles modulaires (materializations incremental/ephemeral/table/view), **macros Jinja** paramétrées, modélisation dimensionnelle (Star Schema, One Big Table) et historisation **SCD Type 2** (snapshots).
- Qualité & CI/CD** : tests dbt génériques et singuliers (unicité, non-null, intégrité référentielle) avec niveaux de sévérité, environnements gérés via workflow Git.

CERTIFICATIONS

Python for Data Science (IBM, 2024) • Data Analysis with Python (IBM, 2023) • SQL and Relational Databases (IBM, 2023)

LANGUES

Français (courant) • Anglais (courant)

ENGAGEMENT ASSOCIATIF & CENTRES D'INTÉRÊT

- Associatif** : Ex-Président (Association SADIM), Vice-Président (Club Data ENSAH), Coordonnateur (Amal Taznagt).
- Sport** : Course à pied et musculation.
- Musique & Lecture** : Apprentissage du saxophone et lecture.