

Yassine OUAZZANI

DATA ENGINEER / FULL-STACK DEVELOPER

✉️ yassineouazzanii@gmail.com

🏡 paris

📞 06 08 09 22 09

Data Engineer spécialisé dans la conception et le déploiement de plateformes d'automatisation IA et de pipelines de données à grande échelle. Expertise en ingestion multi-sources, cloud computing et orchestration de workflows intelligents.

EXPÉRIENCES PROFESSIONNELLES

• Data Engineer – MadelnTracker, Toulouse

Depuis janvier 2025

Projet Phoenix – AI Web Data Extraction Platform/Dpp Marketing

Conception d'un workflow multi-agent (n8n + GPT-4 + MCP) pour transformer des données non structurées (web, vidéos, produits) en jeux de données validés et exploitables.

- Développement d'un pipeline automatisé : ingestion → scraping intelligent → refactoring → validation → génération de Digital Product Passports.
- Automatisation de la diffusion sur les réseaux sociaux pour accroître visibilité et traçabilité.
- Réduction de 95 % du traitement manuel, avec une précision de 85–95 %.

Stack Tech : Git, Github, Python, Docker, Typescript, Multi-agent AI, n8n Cloud, GPT-4 API, MCP Servers, Playwright, Docker .

Projet EasyUDI – Medical Document AI Automation

Mise en place d'un pipeline unifié pour centraliser le traitement de données réglementaires issues de multiples sources (PDF, ERP, bases de données).

- Automatisation pipeline automatisé : ingestion → reconnaissance documentaire (OCR + IA) → validation → prédiction → enregistrement EUDAMED.
- Entraînement de modèles IA génératifs et spécialisés sur des jeux de données médicaux.
- Déploiement sur Google Cloud pour garantir sécurité et scalabilité.

Stack Tech : Python, Google Cloud, OCR, Generative AI, ML training, ERP/DB integration.

Projet LocalizAI - Plateforme d'Analyse Entreprises France POC

Développement d'une plateforme d'ingestion, transformation et analyse des données d'entreprises françaises.

- Extraction via API SIRENE et sources publiques, ETL Python et stockage PostgreSQL.
- Architecture distribuée sous Docker, orchestration Airflow, streaming temps réel via Redpanda.
- Développement des APIs.

Stack Tech : Python (FastAPI, Pandas), SQL, PostgreSQL, Docker, Airflow, Redpanda, Microservices.

TECHNOLOGIES

Cloud Platforms : Google Cloud

Data Engineering : PySpark, Databricks, Kafka, Elasticsearch, Redis

Containerization : Kubernetes, Docker,

DevOps : Git, Github, Azure DevOps, Gitlab, Terraform

Observability, Visualisation & QA : Prometheus, Grafana, notions Airflow / VTOM

CERTIFICATIONS

En cours: Databricks Certified Data Engineer Associate

Databricks Certified Associate Developer for Apache Spark

AWS cloud practitioner

COMPÉTENCES

Cloud Computing : ● ● ● ● ○

SQL : ● ● ● ● ○

Python : ● ● ● ● ○

Shell : ● ● ● ○ ○

Kubernetes : ● ● ● ○ ○

LANGUES

Français Native

Anglais Fluent

› TOEIC : 850/990

DIPLÔMES ET FORMATIONS

- **Master Big data – Mines Paris TECH, Paris**
2024
- **Ingénieur Généraliste – ENSIP, Poitiers**
De septembre 2014 à septembre 2017