

# Mamadou CISSÉ

## Data Engineer & Data Scientist

Spécialisation : Architecture Big Data, Sécurité des Données & IA Générative

Ris Orangis, France • mamadouciiss224e@gmail.com

LinkedIn • Portfolio / GitHub

## Profil

Étudiant en Master 2 **DataScale (Paris-Saclay)**, double compétence **Bio-informatique**. Spécialiste de la chaîne de valeur hybride : de la construction de pipelines Big Data (Spark, Cloud) à la valorisation par l'IA. Recherche un **stage de fin d'études de 6 mois dès Mars 2026**.

## Compétences Techniques

- **Data Engineering & Cloud** : PySpark, Talend, Hadoop, SQL, **Docker**, Databricks, AWS/Azure, Scala.
- **IA & Machine Learning** : Python, **Graph RAG**, Neo4j, **Transformers (HuggingFace)**, NLP, Séries temporelles.
- **Visualisation & Architecture** : Power BI, **API REST (FastAPI/Laravel)**, Microservices, Clean Code.

## Projets Majeurs

- Graph Data Management for Smart Energy Grids** | Neo4j, PySpark, LLM **2025**
  - Conception d'un **Knowledge Graph** unifiant données RTE, Enedis et socio-démographiques (fusion multi-sources).
  - Mise en place d'une architecture **Graph RAG** pour ancrer les LLMs sur des données énergétiques réelles (anti-hallucination).
  - Développement de pipelines ETL pour la simulation de profils de charge horaires et détection de tensions réseau.
- Système OCR Intelligent (Vision NLP)** | Python, TrOCR, Transformers **2024**
  - Développement d'un pipeline de classification et d'extraction automatique sur documents manuscrits bruités.
  - Fine-tuning de modèles **Vision-Transformer** pour l'extraction de métadonnées (Identifiants, notes d'examens).
- Pipeline BI End-to-End** | Talend, Spark, Power BI, Hadoop **2024**
  - Intégration massive de données hétérogènes avec application de règles strictes de **Data Quality**.
  - Création de dashboards décisionnels pour le pilotage de KPIs industriels complexes.

## Expériences Professionnelles

- Stage Data Scientist / Engineer** | OCP Group **Maroc**
  - **Big Data** : Déploiement de pipelines Batch via **Spark** pour le monitoring de capteurs industriels.
  - **Modélisation** : Prédiction de maintenance via réseaux de neurones récurrents (**LSTM**) et Prophet.
  - **Production** : Industrialisation des modèles via API et conteneurisation pour intégration logicielle.
- Stage Analyste de Données Génomiques** | CHU de Tanger **Maroc**
  - Développement d'algorithmes de prétraitement et analyse statistique sur séquences d'ADN complexes.
- Stage Développeur Full Stack** | AFRI Lab **Maroc**

## Formation

- Master 2 DataScale (Ingénierie des Données et Big Data)** **En cours (2026)**  
Université Paris-Saclay (UVSQ) • Focus : Data, Cloud, IA & MLOps.
- Master Bio-informatique et Data Science** | ENSA Tanger **2023**
- Licence Mathématiques et Informatique** | Université Cadi Ayyad **2021**

## Langues & Atouts

- Anglais** : Courant. • **Français** : Langue maternelle.
- Atouts** : Rigueur, excellent esprit d'équipe, grande curiosité et très forte capacité d'adaptation.