

# El Mehdi RAJI

## Data Scientist / Data Engineer | Python

### À propos de moi

Data Engineer avec 3 ans d'expérience dans le traitement du Big Data et la conception de Data Pipelines. Curieux, flexible et autonome, avec une grande aisance relationnelle. À la recherche d'une nouvelle opportunité professionnelle.

### Expérience professionnelle

#### Data Engineer — Capgemini || Stellantis Rennes

Depuis octobre 2022

*Optimiser les performances de la Peugeot 9X8 pour le Championnat du Monde d'Endurance en analysant les données de télémétrie en temps réel pour améliorer la stratégie énergétique :*

- Conception d'un Data Lake structuré sur AWS S3, assurant une architecture scalable et optimisée pour la gestion des données télémétriques.
- Intégration des flux temps réel via Apache Kafka et AWS Glue (PySpark) pour nettoyer, enrichir et agréger les données, garantissant une qualité optimale.
- Développement de pipelines ETL automatisés avec AWS Glue, orchestrés via Apache Airflow, pour un traitement fiable des données.
- Mise en place d'indicateurs de contrôle qualité des données et création de dashboards interactifs sur AWS QuickSight pour le suivi en temps réel.
- Implémentation des rôles et politiques IAM pour sécuriser l'accès aux données (S3, Glue, Redshift), conformément aux normes de sécurité.
- Collaboration avec les équipes métiers et techniques pour formaliser les règles métier, assurer le linéage des données et documenter les processus.

**Technologies :** Python, PySpark, SQL, Apache Kafka, Apache Airflow, AWS (S3, Glue, Redshift, Athena, IAM, QuickSight), Git, Docker, Agile/SAFe.

#### Data Engineer — Capgemini || Stellantis Rennes

De février 2022 à septembre 2022

*Développement d'une plateforme ADAS pour véhicules autonomes, axée sur l'amélioration de la sécurité via l'analyse de données et la gestion avancée de scénarios.*

- Migration des bases MongoDB vers SQL Server avec restructuration de 311 tables pour améliorer la qualité et la disponibilité des données.
- Mise en place de pipelines Kafka pour l'ingestion et la distribution des données capteurs en temps réel, garantissant l'intégrité des flux.
- Orchestration des workflows batch et streaming, incluant automatisation des tests de qualité et alertes proactives.
- Analyse statistique avancée des scénarios ADAS via Python et PySpark pour détecter et corriger les anomalies données.
- Création de dashboards Power BI pour un pilotage précis de la qualité et performance des données ADAS.
- Interface continue avec les équipes métiers pour adapter les processus data aux besoins opérationnels.

**Technologies :** Python, PySpark, SQL Server, MongoDB, Apache Kafka, Databricks, Microsoft Azure, Power BI, Docker, Agile/SAFe.

#### Data Scientist / Data Engineer — Capgemini || Stellantis Rennes

D'octobre 2021 à février 2022

*Développement d'un chatbot intelligent utilisant des techniques NLP pour simplifier la recherche d'informations dans un corpus de 4000 documents.*

- Développement d'un pipeline NLP complet intégrant tokenization, lemmatization, POS tagging et extraction d'entités nommées.
- Optimisation des modèles d'extraction d'informations (n-grams, tf-idf) pour améliorer la pertinence des résultats.
- Création d'algorithmes de recommandation basés sur l'analyse de sentiment pour affiner les réponses du chatbot.
- Intégration d'ElasticSearch pour accélérer l'indexation et la recherche.
- Tests continus et validation des modèles NLP pour garantir la qualité des réponses clients.

**Technologies :** Python (PySpark, Scikit-Learn, Tensorflow, Pytorch), NLP, NLTK, ElasticSearch

### Diplômes et Formation

Ingénieur en Électronique : Machine Learning & Robotique — ENSEIRB-MATMECA France

De 2018 à 2021

✉ elmehdi.raji@outlook.com

🇪🇺 Permis B

📍 France

☎ 07 52 51 49 91

### Certifications

- [Databases and SQL for Data Science](#)
- [Dataiku Core Designer](#)
- [Databricks Fundamentals](#)
- [AgilePM® Foundation](#)

### Atouts

- Travail en équipe
- Rigueur et précision
- Curiosité
- Force de proposition

### Compétences

#### Langages de programmation :

Python, Java, C, C++, Matlab

#### Gestion de bases de données :

SQL, NoSQL (MongoDB, SQL Server)

#### Machine Learning et Analytics :

TensorFlow, PyTorch, Scikit-Learn, Fine-tuning, LLM, RAG

#### Plateformes cloud & Traitement de données :

AWS (Glue, Lambda, S3, QuickSight, IAM), Microsoft Azure, Snowflake, Pandas, PySpark, PowerBI

#### Outils bureautiques :

PowerPoint, Word, Excel

#### Méthodologies de Gestion de Projet :

Méthodologies Agile/SAFe

### Langues

#### Français

Bilingue

#### Anglais

Courant

#### Arabe

Langue maternelle