



# Romain

## Data engineer

Expérience  
**7**  
ans

### Mes motivations

D'une grande curiosité intellectuelle, j'automatise et met en place les systèmes de gestion de la donnée, tout en faisant preuve de pédagogie pour expliquer simplement les solutions aux profils non techniques

06 04 14 92 10

romain.mathonat@gmail.com

### Formation

Docteur en data mining  
- INSA Lyon

Ingénieur en informatique  
- INSA Lyon

### Certifications

Machine learning  
- Coursera  
Deep Learning  
- Coursera  
TDD  
- Michaël Azerhad

### Langues

Français

Anglais

### Compétences clés

#### Techniques :

- **Langages** : Python, Bash, Javascript, HTML/CSS
- **Manipulation des données** : SQL, Polars, Spark, DuckDB, Pandas, Numpy
- **Data storage**: Snowflake, Clickhouse, Elasticsearch, postgresSQL, SQLite
- **Visualisation des données** : Seaborn, Grafana, Streamlit, JS
- **Machine learning** : Sklearn, TPOT, Keras, Pytorch
- **API** : FastAPI, pydantic, Flask
- **Cloud** : Azure, VPS (bare-metal)
- **Orchestration** : Databricks, Airflow
- **Gestion du code**: Git, GithubAction, Gitlab
- **Gestion de l'environnement** : Docker, Poetry, pdm, uv

#### Fonctionnelles :

- **Méthodologie** : Agile Jira/Scrum, Kanban
- **Communication** : Vulgarisation d'un sujet technique
- **Veille technologique** : Synthétiser l'état de l'art d'un sujet

### Expériences professionnelles

#### Michelin – Lead Data Engineer (10 mois)

Mise en place et amélioration de pipelines de données de mobilité

#### Infologic – Data engineer (3 ans 3 mois)

Conception et mise en place du data warehouse, choix des outils et des solutions, collecte et mise à disposition des données

#### Atos – Data engineer (3 ans)

Doctorat fouille de données, développement d'un système d'acquisition de données pour la pousse de plantes en intérieur (start-up)

### Formateur

- **Ecoles** : INSA Lyon
- **Blog personnel**: idem (vulgairedev.fr/about)
- **LinkedIn** : Ecriture d'articles pédagogiques

## Michelin

### Lead Data engineer

10 mois

#### Projet :

Mise en place et amélioration des pipelines de données de mobilité

#### Contexte :

L'entité Michelin Mobility Intelligence (MMI) a pour mission de rendre les infrastructures routières plus sûres, de meilleure qualité, et plus écologique. Pour ce faire, des données de mobilités sont collectées et stockées dans le data warehouse, pour être ensuite transformées et utilisées par divers algorithmes de machine learning pour fournir de la prédiction et de l'analytique au client.

#### Missions :

- Mise en place de data pipelines à haut volume (~100 milliards/mois)
- Amélioration de la CI/CD (performance / facilité d'usage)
- Modernisation du système de dépendances python, refonte du stockage et du build des wheels
- Formation de l'équipe aux bonnes pratiques / outils (TDD, archi hexa, pdm, ruff, wsl, vscode)
- Etat de l'art (data sharing)
- Développement librairie de monitoring de performance interne
- Dashboarding et dataviz

#### Environnement technique :

Snowflake, snowpark (équivalent pyspark), SQL, python, CI/CD, pytest, Microsoft Azure, Databricks, Grafana, Streamlit, Gitlab

## Infologic

### Data engineer / scientist

3 ans 3 mois

#### Projet :

Elaboration et mise en place du data warehouse sur machines bare metal, gestion du cycle de A à Z

#### Contexte :

L'entreprise Infologic est une PME qui fournit un ERP dans le monde agro-alimentaire. Dans ce contexte, de nombreuses données sont disponibles, aussi bien de comportement utilisateur que des données techniques de fonctionnement des serveur. Celles-ci sont sous-exploitées, il a donc fallu créer une architecture permettant de les stocker pour ensuite les utiliser pour divers besoins: maintenance préventive, détection de problèmes d'ergonomie, suivi utilisateur, etc.

#### Missions :

- Data modeling et mise en place du data warehouse
- Formation aux bonnes pratiques pour les data scientists
- Sécurisation et installation des solutions sur les machines (devops)
- Mise en place des datas pipelines
- Développement d'une solution de monitoring in-house
- Etat de l'art sur sujets scientifiques (clustering, LLM)
- Vulgarisation et communication au business

#### Environnement technique :

Python, SQL, Clickhouse, Elasticsearch, Bare metal, notebooks, Jupyter Hub, java

## Atos

### Data engineer

3 ans (mi-temps)

#### Projet :

Système de collecte de données pour une startup de fermes urbaine

#### Contexte :

En collaboration avec une startup: FUL. Le business model était de faire pousser des végétaux dans des environnements de pousse contrôlés: faire pousser des melon au japon devient possible. Ma mission était alors de récolter des données de fonctionnement de cette ferme urbaine, afin de consolider la donnée pour pouvoir entraîner des algorithmes de prédiction de pousse des végétaux

#### Missions :

- Collecte de données (ETL)
- Data visualisation
- Mise en place d'un modèle de prédiction de la consommation électrique
- Participations aux comités de pilotage
- Mise en place d'un modèle "human in the loop" d'amélioration des recettes de pousses de végétaux

#### Environnement technique :

Python, SQL, Linux, pytest, Git, SQLite, Grafana

## Liris

### Doctorat Data Mining

3 ans (mi-temps)

#### Projet :

Recherche, invention d'algorithmes de découvertes de règles interprétables pour données séquentielles, appliquées au e-sport

#### Contexte :

Cette thèse a porté sur la création de nouveaux algorithmes permettant de faire de la fouille de données sur des données séquentielles et temporelles. J'ai appliqué ces travaux à des séquences d'inputs et d'information 3D du jeu "Rocket League". Ceci a mener à diverses publications, ainsi que la possibilité de détecter des comportements utilisateurs jusque là indétectables, qui ont notamment intéressé divers acteurs du milieu (startups, Ubisoft).

#### Missions :

- Etat de l'art scientifique
- Auto-Formation
- Création de nouveaux algorithmes
- Communication et Vulgarisation de mes travaux
- Enseignement

#### Environnement technique :

Python, SQL, Linux, Git, Pytest, notebooks