



Djalil Salah-bey

## Data Engineer | Analytics Engineer

Data Engineer / DataOps orienté fiabilité : CI/CD Docker/Git, tests dbt, monitoring Grafana/OpenLineage. Pipelines clairs, versionnés et mesurables (batch/streaming). Portfolio de projets concrets (code + démos).

### Expériences professionnelles

#### Data Engineer Intern

De mars 2025 à septembre 2025 **PowerUp Technology -EnergyTech** Paris

- Conception et déploiement d'une **Data Platform Databricks Lakehouse** pour la gestion des données batteries IoT (charge/décharge, télémétrie) en collaboration active avec 3 data engineers.
- Implémentation de l'architecture **Médaille (Bronze / Silver / Gold)** avec **Delta Lake** et automatisation des pipelines via **dbt**, gouvernance avec **Unity Catalog**.
- Développement de **pipelines batch et streaming** (Databricks Autoloader, Azure Blob Storage, Kafka) et transformations scalables en **SparkSQL / PySpark**.
- Création de **data marts analytiques** (familles KPIs d'usage, de performance & de sécurité) exposés via **Databricks SQL** et intégrés à **Power BI** (cartes thermiques).
- Préparation de **datasets ML** et suivi du cycle de vie avec **MLflow** en production .
- **Orchestration** via **Databricks Workflows** et intégration **CI/CD** (Git, Azure DevOps, Airflow).
- Gestion des environnements de développement via **Docker** et déploiement via **Azure Kubernetes Services (AKS)**.
- Intégration de bases **SQL & NoSQL** (PostgreSQL, MongoDB, InfluxDB) dans les pipelines de données.
- **Observabilité & qualité** : métriques/logs **Grafana**, **alertings Airflow**, **tests dbt**.

**Stack** : Azure, Databricks, Delta Lake, Spark (PySpark, SparkSQL), dbt, Unity Catalog, Grafana, MLflow, Docker, Kubernetes, PostgreSQL, MongoDB, InfluxDB, Azure Blob, Kafka, Git, CI/CD, Power BI.

#### Data Engineer en alternance

De décembre 2022 à novembre 2024 **Koacher -SportTech** Lyon

- Développement et maintenance d'une **plateforme SaaS** de coaching sportif (web & mobile) au sein d'une équipe de 4 dans le pôle informatique & data analytics.
- Conception et implémentation des **bases de données relationnelles** (PostgreSQL, MySQL).
- **Optimisation SQL** (schémas normalisés, indexation, analyse de plans d'exécution).
- Création d'**APIs REST** sécurisées en **Python/Flask** pour l'intégration front/back (authentification **JWT**, validation des schémas, **rate limiting**).
- Mise en place de **pipelines ETL/ELT** pour centraliser et transformer les données utilisateurs (logs d'entraînement, sessions vidéo) avec **dbt** (modèles, tests, documentation).
- Intégration de solutions **analytics** pour suivre la performance des utilisateurs (**Power BI** : usage, rétention, engagement).
- Gestion du **cycle de développement** avec **Git**, **Docker**, **CI/CD (GitHub Actions)**.
- **Observabilité** et supervision des services avec **Prometheus** (latence API, taux d'erreurs, jobs ETL).

**Stack** : GCP, GKE (Kubernetes), Python, Flask, PostgreSQL, MySQL, SQL, dbt, Power BI, Docker, Git, CI/CD (GitHub Actions), Prometheus.

### Diplômes et Formations

#### Master Ingénierie & Science des données

De 2023 à 2025 **IA SCHOOL** Aix-en-Provence

#### Licence 3 Ingénierie & Science des données

De 2022 à 2023 **IA SCHOOL** Paris

#### BAC +3 Banque & Finance

De 2019 à 2022 **Université de Lille** Lille

✉ salahbeydjalil@gmail.com

🔗 <https://djo-project-portfolio.vercel.app/>

📍 France

☎ +33 6 11 27 91 53

### Langues

Français

Anglais

Arabe

Espagnol

### Centres d'intérêt

#### Sport

Tennis

Basket Ball

Football

#### Voyages

Nouvelles cultures

Tour du monde

#### Tech

Conférences Data

Veille IA

Meetups

### Compétences

- **Data Platforms** : Databricks, Delta Lake, dbt, Unity Catalog
- **Data Engineering** : Spark (PySpark, SparkSQL), ETL/ELT, batch/streaming, Kafka
- **Databases** : PostgreSQL, MySQL, MongoDB, InfluxDB, ClickHouse
- **Cloud & DevOps** : Azure, GCP, Docker, Kubernetes (AKS/GKE), Terraform
- **Orchestration** : Airflow, Databricks Workflows, CI/CD (GitHub Actions, Azure DevOps)
- **Observabilité** : Prometheus, Grafana
- **BI & Apps** : Power BI, Streamlit, APIs Flask
- **ML/MLOps** : MLflow, scikit-learn, TensorFlow, PyTorch