



# Djalil Salah-bey

## Data & Analytics Engineer | En recherche de CDI

Data Engineer passionné par la programmation et l'automatisation des pipelines de données, j'aime concevoir et industrialiser chaque étape pour garantir la fiabilité, la performance et la traçabilité des flux en production.

### Email

salahbeydjalil@gmail.com

### Numéro de téléphone

+33 6 11 27 91 53

### Site web

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

## Langues

### Français

Natif

### Anglais

Courant

## Passions

### Sport

Programmation

Intelligence Artificielle

## Compétences Techniques

- **Cloud & Storage** Azure (Azure Databricks, ADLS Gen2), GCP (BigQuery, GCS)
- **Orchestration** Airflow, Databricks Workflows, Prefect, Cloud Composer
- **Transformation & Modélisation** dbt, SQL, PySpark, Pandas
- **DataOps & CI/CD** GitHub Actions, Terraform (IaC), dbt tests, Marquez/OpenLineage
- **Monitoring & Observabilité** Grafana, Prometheus, Great Expectations
- **Databases / BI** PostgreSQL, MongoDB, DuckDB, Looker Studio, Power BI
- **Languages** Python, PySpark, SQL, Bash
- **DevOps / Conteneurisation** Docker, Kubernetes (minikube, AKS)

## Expériences professionnelles

- **PowerUp Technology** Paris  
De février 2025 à septembre 2025

Freelance Data Engineer

  - Conception et déploiement d'une **plateforme Lakehouse Azure Databricks** centralisant les flux IoT multi-clients (télémétrie rack/module/cellule) pour le monitoring batterie en production.
  - **Ingestion et orchestration automatisées** sous Airflow AKS (SFTP, InfluxDB, PostgreSQL) assurant des pipelines journaliers sans intervention manuelle et traçables via OpenLineage / Marquez.
  - **Nettoyage, normalisation et enrichissement PySpark** des signaux capteurs (température, tension, SOC, SOH) avec contrôle de complétude et cohérence des données.
  - **Modélisation analytique dbt** : calcul des indicateurs de performance, vieillissement, sécurité et usage, avec stratégie incrémentale et qualité vérifiée (tests + CI/CD GitHub Actions).
  - **Supervision et observabilité** complètes via Grafana / Prometheus : suivi des SLA, latence, coûts, et visualisation des heatmaps batterie client en temps réel.
  - **Stack** : Python · Spark · PySpark · SQL · Azure Databricks · Azure Kubernetes Services · Airflow · Azure Data Lake Gen2 · Delta Lake · PostgreSQL · MongoDB · InfluxDB · Grafana · Prometheus · Azure DevOps · Docker · Git
- **Koacher** Lyon  
De septembre 2022 à septembre 2024

Data Engineer

  - **Conception et industrialisation d'un Data Lakehouse GCP** (GCS → BigQuery → dbt) pour la collecte, l'historisation et l'analyse des données issues des capteurs sportifs et des applications mobiles.
  - **Ingestion automatisée** des flux Firebase, Stripe et API capteurs via Cloud Functions / Cloud Composer (Airflow GCP), avec planification quotidienne et gestion d'erreurs centralisée.
  - **Nettoyage et transformation SQL / dbt** (staging → marts) : normalisation, contrôle qualité et versioning SCD Type 2 pour l'analyse long terme des performances utilisateurs.
  - **Modélisation analytique BI** : création d'un schéma en étoile consolidant utilisateurs, capteurs et paiements, garantissant des indicateurs cohérents pour les dashboards métier.
  - **Automatisation et gouvernance DataOps** : tests dbt, documentation, lineage (dbt Docs), CI/CD via GitHub Actions assurant qualité et traçabilité continues.
  - **Stack** : Python · SQL · dbt · Airflow (Cloud Composer) · BigQuery · GCS · Cloud Build · GitHub Actions · Looker Studio · Tableau · Terraform · Docker · Kubernetes · Git

## Projet perso – Septembre 2025

### Démonstration d'un pipeline IoT entièrement automatisé

- Développement d'un **pipeline IoT complet orchestré par Airflow**, collectant chaque jour les flux météo, pollution et trafic routier (APIs OpenWeatherMap, OpenAQ, Transport Data Gouv).
- **Ingestion automatisée** (batch + micro-batch) et transformations **PySpark** en zone Silver : nettoyage, normalisation, partitionnement et historisation optimisée.
- **Modélisation analytique dbt** : schéma en étoile consolidé (météo–pollution–trafic) avec stratégie incrémentale et tests de qualité intégrés.
- **CI/CD, tests et conteneurisation** : Makefile, GitHub Actions, Docker/Kubernetes et tests dbt/Pytest intégrés garantissant la qualité, la portabilité et la stabilité du pipeline.
- **Supervision temps réel** : dashboards Grafana pour SLA et fraîcheur, traçabilité complète via **OpenLineage / Marquez**.

**Stack** : Airflow · Python · PySpark · dbt · DuckDB · Parquet · Grafana · OpenLineage · Marquez · GitHub Actions · Makefile · Docker · Kubernetes · Pytest

**Code** : [github.com/IADJALILProject/iot-smartcity-data-platform](https://github.com/IADJALILProject/iot-smartcity-data-platform)

## Diplômes et formations

- **IA School** — Diplôme RNCP Niveau 7 Expert Data & Ingénierie de l'Intelligence Artificielle (2022-2024)
- **IAE de Lille** — Master Système d'information & management de la donnée (2021-2022)
- **Université de Lille** — Licence Banque, Finance (2018-2021)