



## 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, déployer et industrialiser chaque étape pour garantir la fiabilité, la performance, la traçabilité et la disponibilité 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

## Compétences Techniques

### Cloud & Data Platform

- Azure (Databricks, ADLS Gen2), GCP (BigQuery, GCS)

### Orchestration & Scheduling

- Airflow (AKS, GKE), Databricks Jobs/Workflows

### Modélisation & Transformation

- dbt (modèles métiers, tests, documentation)
- SQL, PySpark, Pandas

### DataOps & Qualité

- GitHub Actions (CI/CD), Terraform (IaC)
- dbt tests, Pytest, Great Expectations

### Monitoring & Observabilité

- Grafana, Prometheus, OpenLineage, Marquez

### Databases & BI

- PostgreSQL, MongoDB, DuckDB, ClickHouse
- Looker Studio, Power BI

### Développement &

#### Conteneurisation

- Python, SQL, Bash
- Docker, Kubernetes (minikube, AKS)

## Diplômes et formations

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

## Expériences professionnelles

- PowerUp Technology -EnergyTech** Paris  
De février 2025 à octobre 2025  
Freelance Data Engineer
  - Mise en place de pipelines d'ingestion orchestrés par Airflow (AKS)**, assurant la collecte quotidienne des fichiers clients via SFTP et leur historisation traçable dans **ADLS Gen2 (zones Bronze → Silver → Gold)**.
  - Développement et industrialisation de traitements PySpark sur Azure Databricks** pour le **nettoyage, la validation et la consolidation des données clients**, assurant le calcul massif des indicateurs de performance, sécurité et vieillissement des batteries.
  - Conception et création d'une base PostgreSQL pour le stockage des indicateurs consolidés** par client, utilisée comme source principale des dashboards clients.
  - Implémentation d'une base ClickHouse (OLAP)** pour permettre des visualisations interactives, hiérarchiques et temps réel (heatmaps batterie, drill-down client → site → rack → module) au sein des tableaux de bord.
  - Développement et structuration d'une base MongoDB pour le back-office interne**, gérant les métadonnées, configurations et entités métiers, avec conception des collections, index et schéma d'accès optimisé.
  - Mise en place d'une supervision complète** sous Grafana, avec alerting et traçabilité des flux Airflow (logs, SLA, lineage).
  - Intégration d'une chaîne CI/CD Azure DevOps** automatisant les tests, la validation et le déploiement des DAGs Airflow, notebooks Databricks et schémas PostgreSQL/MongoDB/ClickHouse sur les environnements dev → préprod → prod.
  - Architecture entièrement automatisée** (Airflow → Databricks → CI/CD) garantissant un fonctionnement autonome sans intervention manuelle et un taux de réussite supérieur à **99 %** sur les exécutions Airflow.
  - Contribution à la qualité logicielle et au travail collaboratif** : structuration des dépôts Git (arborescence, conventions, modularisation), revue de code, refactoring continu et documentation technique centralisée.

**Stack :** SQL · Python · PySpark · Azure Databricks · ADLS Gen2 · pytest · Airflow · Azure Kubernetes Services · PostgreSQL · ClickHouse · MongoDB · Azure DevOps · Grafana · Git · Docker · Kubernetes.
- Koacher -Sportech** Lyon  
De septembre 2022 à novembre 2024  
Data Engineer
  - Conception et déploiement d'une plateforme Data GCP** centralisant les données de l'application Koacher (utilisateurs, séances, performances, transactions Stripe, analytics Firebase) au sein d'un Data Lakehouse BigQuery.
  - Déploiement et administration d'Airflow sur GKE**, orchestrant les flux d'ingestion et de transformation quotidiens, avec gestion des dépendances, logs et alertes pour garantir la fiabilité du pipeline.
  - Conception et implémentation de modèles dbt** pour la transformation et la modélisation analytique (staging, marts), avec tests de qualité intégrés (freshness, uniqueness, relationships) et documentation générée automatiquement.
  - Mise en place d'une chaîne CI/CD GitHub Actions** automatisant les tests dbt, le linting SQL et le déploiement des modèles sur les environnements dev → prod.
  - Collaboration avec les équipes produit et marketing** pour alimenter les dashboards Data Studio / Looker Studio, garantissant la cohérence, la fraîcheur et la fiabilité des indicateurs de suivi utilisateurs et business.
  - Contribution à la qualité logicielle** : structuration du repository dbt, refactoring des modèles, normalisation des conventions et documentation technique.

**Stack :** SQL · Python · GCP · BigQuery · Cloud Storage · Airflow · GKE · dbt · GitHub Actions · Data Studio · Git · Docker · Kubernetes

## Projet perso – Septembre 2025

Pipeline IoT autonome orchestré par Airflow, collectant et historisant les flux publics OpenWeatherMap, OpenAQ et Transport Data Gouv pour analyser l'impact du trafic routier sur la qualité de l'air à Paris.

**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)