



# Djalil Salah-bey

## Data Engineer – Pipelines, Lakehouse & DataOps (Azure / GCP)

Data Engineer passionné par la construction de plateformes de données fiables et évolutives. J'aime concevoir des pipelines clairs, documentés et automatisés pour garantir la qualité et la disponibilité des données. Habitué aux environnements Azure Databricks et GCP BigQuery, j'applique au quotidien une approche DataOps basée sur la rigueur, la collaboration et l'amélioration continue.

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

**Email**  
salahbeydjalil@gmail.com

**Numéro de téléphone**  
+33 6 11 27 91 53

### Langues

**Français**  
Natif

**Anglais**  
Courant

### Passions

Sport  
Programmation  
Intelligence Artificielle

### Diplômes et formations

- **Microsoft Azure DP-203** – Certification Data Engineer Associate (en cours)
- **IA School** — Diplôme RNCP Niveau 7 Expert Data & Ingénierie de l'Intelligence Artificielle (2023-2025)
- **IA School** — Bachelor Expert Data & Ingénierie de l'Intelligence Artificielle (2022-2023)
- **Université de Lille** — Licence Banque, Finance (2019-2022)

### Compétences Techniques

**Langages** : Python, R, SQL, Shell  
**Cloud & Plateformes** : Azure (Databricks, ADLS, DevOps), GCP (BigQuery, GCS, Cloud Composer)  
**Orchestration & Modélisation** : Airflow, dbt Core, Dataflow, Delta Lake  
**CI/CD & Observabilité** : Cloud Build, Prometheus, Grafana, Cloud Monitoring  
**Bases & BI** : PostgreSQL, MongoDB, InfluxDB, Power BI, Looker Studio  
**Big Data** : Hadoop, Spark  
**IA** : MLFlow, Sickit-Learn, Tensorflow, Keras, OpenCV, SciPy

### Expériences professionnelles

- PowerUp Technology** Paris  
Data Engineer De février 2025 à octobre 2025
  - Pipelines batch **SFTP → ADLS → Delta Lake (PySpark)** : ingestion, contrôles, logs et traçabilité clients.
  - **Notebooks Databricks** : agrégations, enrichissements et indicateurs de performance batterie.
  - **Airflow (AKS)** : orchestration des jobs Databricks, dépendances, retries et alertes SLA.
  - **Tests PySpark** : validation des formats, outliers et cohérences temporelles en production.
  - **CI/CD Azure DevOps** : packaging, déploiement et documentation automatisée des notebooks.
  - **Prometheus + Grafana** : supervision Databricks/Airflow, suivi SLA et performances.
  - **Data Science & IoT** : préparation et fiabilisation des jeux de données prédictifs.
  - **Power BI** : visualisation des indicateurs batterie (charge, température, alertes).
  - **DataOps / MLOps** : CI/CD Databricks, versioning des artefacts, suivi modèles prod.
  - **Intégration applicative** : KPI (PostgreSQL), configurations (MongoDB), mesures (InfluxDB).
  - **Datas marts** : Modélisation sous ClickHouse pour l'intégration incrémentale des données pour les Heatmaps 3D.

**Stack** : Python · PySpark · Azure Databricks · Delta Lake · ADLS Gen2 · Airflow · Azure DevOps · Prometheus · Grafana · Docker · PostgreSQL · MongoDB · InfluxDB · ClickHouse · Power BI.
- Koacher** Lyon  
Data & Analytics Engineer De décembre 2022 à décembre 2024
  - Plateforme analytique **GCP** centralisant **Firebase, Stripe** et APIs sportives dans **BigQuery**.
  - **Airflow (Cloud Composer) & Dataflow (Beam)** : ingestion quotidienne vers **GCS**, reprise sur incident.
  - **dbt Core + BigQuery (Kimball SCD2)** : modélisation des dimensions, faits sessions et paiements.
  - **Tests dbt** : cohérence paiements-utilisateurs, outliers et doublons sur modèles analytiques.
  - **Airflow avancé** : DAGs dynamiques, dépendances, backfills et alerting SLA.
  - **dbt Docs + BigQuery metadata** : documentation et lineage automatisés.
  - **Cloud Monitoring / Logging** : supervision des pipelines et contrôle des coûts BigQuery.
  - **Cloud Build CI/CD** : tests, génération docs et déploiement multi-environnements.
  - **Optimisation BigQuery** : partitioning, clustering et tuning SQL (−35 % coûts, −40 % latence).
  - **Looker Studio / Power BI** : dashboards de suivi athlètes et indicateurs de performance.
  - **DataOps / Analytics Engineering** : industrialisation pipelines dbt et observabilité continue.

**Stack** : Python · SQL · dbt Core · BigQuery · GCS · Airflow · Firebase · Stripe API · Cloud Build · Cloud Monitoring · Cloud Logging · Docker · Looker Studio.
- Muli By Easy Planet** Paris  
Referent ETL & Flux De janvier 2022 à juillet 2022
  - Développement de **scripts Python** pour automatiser les chargements et contrôles de données.
  - Conception et maintenance de **flux ETL Talend** pour alimenter les bases **PostgreSQL**.
  - Intégration et planification des traitements dans **Airflow** pour un enchaînement fiable des jobs.
  - Contribution à la mise en **production et surveillance** des pipelines de données.
  - **Renforcement des bonnes pratiques DataOps** : tests, documentation et suivi de la qualité des flux.

**Stack** : Talend · Python · SQL · PostgreSQL · Airflow · Hadoop · Spark