



# Djalil Salah-bey

## Data Engineer | Analytics Engineer

Spécialisé en traitement Big Data (Python / PySpark), modélisation analytique dbt, orchestration Airflow et architectures Lakehouse. Forte expérience en production : CI/CD, tests automatisés, optimisation des jobs, observabilité, robustesse des pipelines et engineering de qualité.

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 & Plateformes**  
Azure : Databricks, ADLS Gen2, Azure DevOps, AKS, Key Vault  
Compute : Spark / PySpark  
Warehouses : ClickHouse, SQL Server, PostgreSQL
- **Orchestration**  
Airflow (AKS)  
Databricks Workflows
- **Modélisation & Transformation**  
dbt (staging, dims, facts, marts)  
Data Lakehouse / Delta Lake SCD, tests, lineage
- **DevOps / DataOps**  
Azure DevOps (YAML)  
Tests automatisés (dbt, Python)  
Packaging & déploiement automatisé  
Monitoring & observabilité (logs, SLA, metrics)
- **Langages**  
Python, PySpark, SQL

### Diplômes et formations

- **IA School —Paris Master Ingénierie des données & de l'intelligence artificielle** (RNCP Niveau 7) ( 2023-2025)
- **IA School —Paris Licence 3 Ingénierie des données & de l'intelligence artificielle** (2022-2023)
- **Université de Lille — Licence Developpeur Informatique** (2019-2022)

### Expériences professionnelles

- PowerUp Technology**  
Data Engineer

Paris  
D'octobre 2024 à octobre 2025

**Conçu et industrialisé des pipelines PySpark sur Azure Databricks** : ingestion quotidienne, normalisation, transformations incrémentales Delta et alimentation fiable des zones Bronze / Silver / Gold.

**Orchestré les traitements des séries temporelles avec Databricks Workflows** : gestion des dépendances, planification, monitoring des runs et optimisation de la stabilité des jobs en production.

**Structuré et administré le Data Lake ADLS Gen2** : organisation logique des zones, partitionnement, formats optimisés (Parquet / Delta) et gestion des accès pour garantir la qualité et la sécurité des données.

**Exploité et fiabilisé les pipelines batch sous Airflow (AKS)** : supervision des DAGs, gestion des workers Kubernetes, résolution proactive des incidents et amélioration continue de la disponibilité.

**Mis en place des pipelines CI/CD via Azure DevOps** : automatisation des tests, packaging PySpark, déploiements contrôlés et bonnes pratiques Git pour sécuriser le cycle de livraison.

**Construit des dashboards de supervision sur Grafana** : suivi de la qualité des pipelines, des SLA journaliers et des indicateurs de santé des workflows.

**Structuré l'ensemble des pipelines en repositories dédiés** (ingestion, transformations PySpark, orchestration Airflow/Workflows) pour maximiser la lisibilité du code, la maintenabilité et la robustesse du cycle de vie applicatif.

  - **Stack** : Python · Azure Databricks · PySpark · Airflow sur AKS · ADLS Gen2 · Azure DevOps · Delta Lake · Grafana
- Koacher**  
Data Engineer · Contrat d'alternance

Lyon  
De septembre 2022 à août 2024

**Développé des pipelines de données de bout en bout dans Azure Databricks** : ingestion depuis APIs, nettoyage, normalisation, transformations PySpark et alimentation des tables Delta pour les usages analytiques.

**Automatisé les traitements via Airflow (AKS)** : planification quotidienne, gestion des dépendances, monitoring des DAGs et résolution des incidents courants en production.

**Modélisé les données métier avec dbt et Databricks Warehouse** : tables de faits/dimensions, tests de qualité intégrés, documentation et optimisation des requêtes SQL pour les dashboards produits.

**Construit et maintenu des data marts analytiques (schéma en étoile)** : faits, dimensions, règles métier, métriques consolidées et exposition de KPIs fiables aux équipes produit.

**Mis en place des pipelines Delta Live Tables (DLT)** : ingestion structurée, logique incrémentale, qualité intégrée et historisation fiable des données sportives et de capteurs.

**Intégré les données Stripe pour la gestion des abonnements** : ingestion automatisée, historisation des événements, structuration analytique et exposition des métriques produit.

**Construit des tableaux de bord analytiques et opérationnels sous Prometheus** : KPIs d'usage, rétention, performance sportive et monitoring de la fraîcheur des données.

**Géré le cycle de vie du code via Azure DevOps** : organisation en repositories dédiés, revue de code, tests automatisés et déploiements contrôlés des traitements Databricks.

  - **Stack** : Azure · Databricks · PySpark · Airflow sur AKS · Delta Lake · Databricks Warehouse · dbt · Delta Live Tables · Firebase · Stripe · Prometheus · Python · SQL

### Projet Personnel 2025

- Environnement** : Apache Airflow · PySpark · dbt · DuckDB · Python · APIs REST · Parquet · Docker · Kubernetes · GitHub Actions · Grafana
- Projet démonstratif d'un pipeline Data automatisé** exécuté chaque nuit via Apache Airflow ( APIs Publiques -> Bronze -> Silver -> Gold -> Grafana).
- **Collecte automatisée** des APIs publiques (météo, pollution, trafic) et stockage des données brutes au format Parquet.
  - **Transformations PySpark** : nettoyage des séries temporelles, partitionnement logique et traitements incrémentaux.
  - **Stockage analytique DuckDB** : moteur local rapide pour l'exploration SQL, les tests de qualité, les transformations et la validation intermédiaire des datasets.
  - **Modélisation analytique dbt** (staging → marts) pour produire des indicateurs consolidés environnementaux.
  - **Monitoring via Grafana** : dashboards de fraîcheur, SLA et alertes email en cas d'échec.
  - **CI/CD GitHub Actions** : tests automatisés, build, packaging Docker et déploiement Kubernetes.
- Projet open source** : [github.com/IADJALILProject/iot-smartcity-data-platform](https://github.com/IADJALILProject/iot-smartcity-data-platform)