



# Djalil Salah-bey

## Data Engineer | Developpeur Data

Expérience confirmée dans la mise en production et la supervision de pipelines Databricks avec une approche centrée sur la fiabilité, la scalabilité et l'automatisation.  
Capacité à maintenir des environnements data, à résoudre les incidents de production et à améliorer continuellement la performance et les coûts.

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

### Passion

Programmation

Basket Ball

Football

Intelligence Artificielle

### Compétences Techniques

- Langages : Python, SQL, PySpark
- Data Platforms : Databricks, GCP, Azure, BigQuery
- Orchestration & CI/CD : Airflow, GitHub Actions, Docker, Terraform
- Transformation : dbt, SQL analytique, SCD, tests & docs
- Stockage/Formats : Delta Lake, Parquet, JSON/Avro, ADLS/S3-like
- Streaming/Intégration : Kafka, Auto Loader, Fivetran, CDC, Talend
- BI & Gouvernance : Power BI, Unity Catalog, RBAC, OpenLineage, Grafana, Marquez, Prometheus

### Soft Skills

Sens du détail  
Collaboration  
Communication technique  
Esprit d'automatisation  
Amélioration continue

### Expériences professionnelles

- PowerUp Technology

Data Engineer ( Freelance )

Paris

De mars 2025 à octobre 2025

Conception et exploitation d'une plateforme de données sur **Azure Databricks** dédiée au suivi et à la prévision des performances de batteries industrielles connectées (BESS).

- Intégration et synchronisation des différentes sources de données (API, SFTP, MongoDB) au sein du Data Lake ADLS Gen2 afin d'assurer la cohérence et la fiabilité des flux.
  - Développement de traitements Python & PySpark pour la préparation et la qualité des données : complétion, validation, pré-traitement, filtrage et formatage.
  - Calcul et mise à jour quotidienne des indicateurs clés (SOH, SOC, cycles, températures, alertes) utilisés pour le pilotage et la maintenance prédictive.
  - Stockage et mise à disposition des KPI dans PostgreSQL et InfluxDB pour les tableaux de bord et le suivi opérationnel.
  - Collaboration étroite avec l'équipe Data Science pour la fourniture de données fiabilisées aux modèles de prédiction.
  - Automatisation en batch des flux ETL/ELT via Airflow et Databricks Jobs avec chargements incrémentals, alertes sur Slack et contrôles automatisés.
  - Mise en place d'un processus de supervision et de résolution d'incidents pour garantir la fiabilité et la continuité des flux de production.
  - Intégration continue sous Azure DevOps CI/CD avec gestion des dépendances et exécution des scripts via Databricks CLI.

Stack principale :  
Azure Databricks · Python · PySpark · SQL · ADLS Gen2 · Airflow · MongoDB · PostgreSQL · InfluxDB · Azure DevOps · Grafana · Prometheus · Docker · Kubernetes · APIs · SFTP
- Koacher

Data Engineer ( Alternance )

Lyon

De décembre 2022 à décembre 2024

Contribution à la conception et au maintien d'une plateforme analytique sur **Databricks (GCP)** destinée à mesurer la performance et l'engagement des athlètes au sein d'une application de coaching digital.

- Collecte, ingestion et intégration automatisée des données issues des API mobiles, CRM et Stripe ( API Paiement) vers le Data Lake GCS orchestré par Airflow.
  - Développement de pipelines Python, PySpark et SQL pour la préparation, la standardisation et la fiabilisation des données utilisateurs, sessions et programmes d'entraînement.
  - Conception et modélisation de la couche analytique sous Data Build Tools avec gestion des dépendances et documentation automatisée.
  - Calcul et mise à jour régulière des indicateurs métiers : rétention, activation, fréquence d'usage, churn et progression, utilisés pour piloter la stratégie produit.
  - Stockage et exposition des jeux de données consolidés dans BigQuery, visualisation des analyses via Power BI et Streamlit pour les équipes produit et direction.
  - Orchestration et automatisation des traitements via Airflow et GitHub Actions, incluant tests dbt, alertes sur Slack et suivi des performances sur Grafana.
  - Collaboration étroite avec les équipes Data Science et Produit pour la fourniture de données prêtes à l'emploi et l'enrichissement des modèles prédictifs.

Stack principale :  
Databricks (GCP) · PySpark · SQL · dbt · Airflow · Delta Lake · BigQuery · Power BI · Streamlit · GitHub Actions · Grafana · Terraform · Firebase · Stripe · APIs

### Diplômes et formations

- IA School — Master Expert Data & Ingénierie de l'Intelligence Artificielle (2023-2025)
- IA School — Bachelor 3 Expert Data & Ingénierie de l'Intelligence Artificielle (2022-2023)
- Université de Lille — Licence Banque, Finance & Analyse Quantitative (2019-2022)