



Djaliil Salah-bey

Data Engineer | Spécialisé Databricks, Airflow, DBT & CI/CD

Data Engineer autonome, je conçois et automatise des pipelines de données clairs, robustes et maintenables. J'interviens sur tout le cycle de vie de la donnée — ingestion, transformation, qualité et orchestration — avec un souci constant de rigueur et d'efficacité.

Passionné par les environnements cloud modernes, je cherche à faire évoluer des projets data à fort impact.

Email
salahbeydjaliil@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

Compétences Techniques

Traitement & Plateforme Data

Databricks · Spark · Delta Lake · ClickHouse · Kafka · Airflow · dbt

Stockage & Bases de données

Azure Blob Storage · PostgreSQL · SQL Server · MongoDB · InfluxDB

Orchestration & Automatisation

Airflow · GitHub Actions · Azure DevOps · Docker

Qualité & Gouvernance

dbt tests · Unity Catalog · Great Expectations

Monitoring & Supervision

Grafana · Prometheus · Azure Monitor

Visualisation & Analyse

Power BI · Streamlit

Langages &

Environnements

Python · SQL · YAML · Bash · Azure · GCP · Linux (WSL2)

Soft Skills

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

Expériences professionnelles

- PowerUp Technology

Paris

Data Engineer (Freelance)

De mars 2025 à octobre 2025

Ingestion & Intégration IoT

-Collecter et centraliser les flux télémetriques des batteries BESS via **Kafka**, **API REST** et **Azure Blob Storage**.
-Construire des pipelines d'ingestion **batch et streaming** sous **Databricks** pour alimenter le Lakehouse.

Transformation & Modélisation

-Nettoyer, enrichir et structurer les données avec **Spark / PySpark** selon l'architecture **Bronze → Silver → Gold**.
-Concevoir des tables métiers normalisées (énergie, température, charge, décharge, cycle de vie).

Orchestration & Automatisation

-Développer et planifier les workflows **Airflow** (initdb, ingestion, build_fact, build_mart) avec gestion des dépendances et SLA.
-Mettre en place des alertes et notifications automatiques pour garantir la fiabilité des traitements.

Qualité & Gouvernance

-Implémenter des tests de fraîcheur et de cohérence avec **dbt** et **Great Expectations**.
-Assurer la traçabilité et le lineage des données via **Unity Catalog** et **OpenLineage**.

Optimisation & Monitoring

-Optimiser les performances des jobs Spark (partitioning, Z-Order, caching, tuning des -clusters).
-Superviser les métriques de performance et de coûts via **Grafana**, **Prometheus** et **Azure Monitor**.

Collaboration & CI/CD

-Intégrer les jobs dans la chaîne **Azure DevOps CI/CD**, versionner les notebooks et automatiser les déploiements.
-Collaborer avec les Data Scientists pour la création et la mise à jour des **feature tables MLflow**.

Stack : Databricks · Spark · Delta Lake · Kafka · Airflow · dbt · Python · SQL · ClickHouse · Azure Blob Storage · Grafana · Prometheus · MLflow

Résultats : Observabilité temps réel complète des pipelines Databricks.

●

Koacher

Lyon

Data Engineer (Alternance)

De décembre 2022 à décembre 2024

Ingestion & Intégration

-Centraliser les données issues des **applications web & mobile** (API, PostgreSQL, Firebase, fichiers).
-Construire les pipelines d'ingestion **batch** et de synchronisation sous **Airflow & Python**.

Transformation & Structuration

-Nettoyer, enrichir et préparer les datasets utilisateurs, sessions et événements d'entraînement.
-Mettre en place une architecture de traitement **ETL & ELT** intégrée au **Data Warehouse analytique**.

Automatisation & Orchestration

-Automatiser les workflows analytiques via **Airflow** et gérer les dépendances entre modèles.

Qualité & Traçabilité

-Implémenter des **tests unitaires et de fraîcheur** pour garantir la cohérence des données utilisateurs.
-Assurer le suivi des exécutions et des sources dans un référentiel central (metadata lineage).

Collaboration & CI/CD

-Intégrer les pipelines dans la chaîne **GitHub Actions** pour validation et déploiement continu.
-Collaborer avec les équipes Produit et BI pour enrichir les flux métier (utilisateurs, coachs, sessions).

Stack : Python · SQL · Airflow · dbt · PostgreSQL · BigQuery · Firebase · Docker · Power BI · Grafana

Résultats : Data pipelines dbt & Airflow industrialisés en CI/CD.

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)