



Djalil Salah-bey

Data Engineer & Analytics Engineer

Ingénieur Data spécialisé dans la conception, l'automatisation et la supervision de plateformes de données en production.
J'assure la fiabilité, la scalabilité et la traçabilité des flux de bout en bout — de l'ingestion à l'exposition analytique.

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

- **Langages & frameworks**
Python (pandas, PySpark), SQL analytique, Bash
- **Data Platform**
Databricks (Azure / GCP), Delta Lake, BigQuery, ADLS Gen2
- **Orchestration & CI/CD**
Airflow, Databricks Jobs, GitHub Actions, Azure DevOps
- **Transformation & modélisation**
dbt (tests, docs, SCD2), Kimball / Data Vault
- **Streaming & ingestion**
Kafka, Auto Loader, Fivetran, APIs REST, SFTP
- **Stockage & bases**
PostgreSQL, MongoDB, InfluxDB, Parquet / JSON / Avro
- **Monitoring & gouvernance**
Grafana, Prometheus, Marquez / OpenLineage, Unity Catalog, RBAC
- **Infra & DevOps**
Docker, Terraform, Kubernetes

Soft Skills

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
Mission de production pour le déploiement et la supervision d'une **plateforme Azure Databricks** intégrée à la **roadmap produit PowerUp**, dédiée à l'analyse et la maintenance prédictive d'un **parc de plus de 20 000 batteries industrielles connectées (BESS)**.
 - **Architecture Lakehouse** : structuration du **Data Lake ADLS Gen2** en couches *Bronze / Silver / Gold*, ingestion multi-sources (**API, MongoDB, SFTP**) et gestion automatisée des schémas évolutifs.
 - **Industrialisation Databricks** : mise en production de notebooks **PySpark** (data completion, validation, prétraitement, calcul de KPI) orchestrés sous **Databricks Jobs** avec politiques d'alerting et de reprise automatique.
 - **Optimisation des coûts** : ajustement du **cluster autoscaling**, compression **Delta Lake** et utilisation du **cache adaptatif**, permettant une **réduction de 25 % des coûts de calcul mensuels**.
 - **Supervision & observabilité** : déploiement d'un **monitoring Grafana + Prometheus** et implémentation du **lineage OpenLineage / Marquez** pour assurer la traçabilité complète des pipelines.
 - **CI/CD Azure DevOps** : automatisation des tests, packaging et déploiements **Databricks** via pipelines **YAML**, avec gestion des dépendances et validations continues.
 - **Collaboration Data Science** : fourniture de **datasets fiabilisés et historisés** pour l'entraînement des modèles prédictifs (SOH, SOC, cycles, température), avec suivi des versions sous **MLflow**.
 - **Support métier & opérationnel** : exposition des indicateurs dans **InfluxDB** et **Power BI**, supervision temps réel dans **Grafana**, permettant le suivi proactif des alertes et de la performance des systèmes.

Mise en production réussie du pipeline **BESS Heatmap**, garantissant une **fraîcheur de données inférieure à 1 h** et contribuant à une **réduction de 15 % des interventions non planifiées** sur les sites clients.
- **Koacher** Lyon
Data Engineer (Alternance) De décembre 2022 à décembre 2024
mise en production d'une **plateforme analytique sous BigQuery (GCP)** destinée à mesurer la performance, l'engagement et la rétention des athlètes sur une application de coaching digital.
 - **Architecture analytique** : conception du modèle **staging → intermediate → marts** sous **dbt**, avec tests, documentation automatique et historisation SCD2.
 - **Ingestion & fiabilisation** : intégration automatisée des flux **API mobiles, CRM, Firebase, Stripe** vers **GCS**, orchestrée par **Airflow** avec gestion des dépendances et contrôles de complétude.
 - **Transformation & métriques business** : calcul des indicateurs clés (rétention, churn, activation, fréquence d'usage, progression) utilisés pour piloter la stratégie produit.
 - **Orchestration CI/CD** : automatisation des pipelines via **Airflow + GitHub Actions**, exécution des tests **dbt** et déploiement continu avec **Terraform**.
 - **Supervision & qualité** : suivi des jobs et SLA dans **Grafana**, alerting automatique sur Slack, traçabilité assurée via **OpenLineage / Marquez**.
 - **Exposition & valorisation** : publication des jeux de données consolidés dans **BigQuery**, visualisation dans **Power BI** et **Streamlit** pour les équipes Produit & Direction.
 - **Gouvernance & collaboration** : structuration des conventions de nommage, des modèles et des accès **RBAC** sous **Unity Catalog**.

Industrialisation complète du pipeline analytique Koacher, fiabilisation de 100 % des indicateurs critiques et amélioration de 25 % de la qualité des données consommées par les dashboards décisionnels.

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)**