



Djalil Salah-bey

Data Engineer | Analytics Engineer

- Ingénieur Data spécialisé dans la conception et l'industrialisation de plateformes analytiques.
- J'assure la collecte, le traitement et la fiabilité des données au sein d'environnements cloud modernes.
- Mon travail vise à rendre la donnée accessible, gouvernée et exploitable pour la prise de décision.
- Portfolio : <https://djo-project-portfolio.vercel.app/>

Expériences professionnelles

Data Engineer Databricks (Freelance 8 mois)

De mars 2025 à octobre 2025 **PowerUp Technology** Paris

- Conception et maintenance de pipelines sur **Azure Databricks** pour le suivi de batteries industrielles.
- Mise en place d'une architecture **Delta Lake (Bronze → Silver → Gold)** robuste et scalable.
- Développement de traitements **batch et streaming** connectés à **Kafka** et **Azure Blob Storage**.
- Exploitation de **Spark SQL**, **MLflow** et des **notebooks collaboratifs Databricks**.
- Intégration et normalisation de données **multi-sources** : IoT, API, bases relationnelles.
- Exploitation des fonctionnalités avancées de **Delta Lake** : **time travel**, **merge**, **versioning** et **vacuum**.
- Collaboration avec les Data Scientists pour **industrialiser les modèles ML** et gérer les **feature tables**.
- Automatisation et supervision des workflows via **Airflow**, **Docker**, **Azure DevOps**.
- Mise en place d'un **monitoring applicatif** pour le **suivi des jobs et des coûts Databricks**.
- Application des pratiques **DataOps & CI/CD** : versioning, tests, déploiements automatisés.
- Mise en œuvre du **contrôle qualité, sécurité et traçabilité** (dbt, Unity Catalog, Key Vault).
- Optimisation de la **scalabilité et des performances Spark** sur cluster Databricks.
- **Gestion des accès et lineage centralisé via Unity Catalog et Azure Key Vault**.
- Participation à la définition des **stratégies de scaling et de gouvernance Azure Databricks**.

- **Résultats** : flux unifiés, jobs industrialisés, supervision temps réel, – 40 % de latence de traitement.
- **Stack** : **Azure, Databricks, Delta Lake, Spark/PySpark, Airflow, dbt, Kafka, ADLS, MLflow, Unity Catalog, Azure DevOps, Docker, Grafana, Prometheus, MongoDB, Clickhouse.**

Data Engineer DBT (Alternance 2 ans)

De décembre 2022 à décembre 2024 **Koacher** Lyon

- Conception et déploiement du **Data Warehouse analytique central** de la plateforme Koacher.
- Modélisation des données **staging → intermediate → marts** sous **dbt Cloud**.
- Développement de modèles **SQL testés, documentés et versionnés**.
- Intégration et orchestration des pipelines via **Airflow**.
- Mise en place d'une **chaîne CI/CD** avec **GitHub Actions** pour tests et déploiements automatisés.
- Suivi de la **fraîcheur des sources** et du **lineage automatique** via dbt Cloud.
- Utilisation de **macros Jinja** et de **packages dbt-utils / audit-helper** pour standardiser les modèles.
- Création d'un **catalogue de métriques business** : rétention, engagement, performance sportive.
- Gestion de la **qualité des données** : tests dbt, validation, fraîcheur et cohérence des sources.
- Intégration des **tests dbt** dans la chaîne CI/CD pour validation avant déploiement.
- Supervision des workflows et alertes via **Grafana** et **Prometheus**.
- Contribution à la **gouvernance analytique** et à la **documentation automatisée** (linéage, docs)
- Collaboration avec les équipes produit et data science pour le **reporting et les datasets ML**.
- **Optimisation des modèles SQL** et **suivi** des performances analytiques en temps réel.

- **Résultats** : standardisation des modèles analytiques, réduction de 60 % du temps de requête et adoption totale du framework dbt en production.
- **Stack** : **dbt Cloud, Jinja, Airflow, PostgreSQL, BigQuery, Python, Docker, Grafana, Prometheus, GitHub Actions, Great Expectations, Marquez, OpenLineage, Power BI.**

Data Engineer Talend (Stage 6 mois)

De janvier 2022 à juillet 2022 **Mutli By EasyPlanet** Paris

- Conception et maintenance de **workflows ETL** sous **Talend, Python** et **SQL**.
- Intégration de données multi-sources **ERP & CRM** pour la finance et la logistique dans PostgreSQL.
- Automatisation des traitements récurrents et planification **incrémental** via **Airflow**.
- Développement de **routines Talend** avec gestion des dépendances et des logs.
- Mise en place de **contrôles qualité** et de mécanismes de **traçabilité** des flux.
- Déploiement de tableaux de bord **Power BI** pour le reporting et la visualisation.
- Contribution à la **migration des flux ETL** vers une **architecture Cloud** (BigQuery)
- **Stack** : **Talend DI/ESB, Airflow, SQL Server, PostgreSQL, Python, Power BI, Git, BigQuery.**

Diplômes

IA School — Master Expert Data & Ingénierie de l'Intelligence Artificielle

De 2022 à avril 2025 **IA SCHOOL** Paris

IAE Lille — Master Systèmes d'Information & Management de la Donnée

De 2021 à 2022 **IAE Lille** Lille

Université de Lille — Licence Spécialisé en Banque et Finance

De 2018 à 2021 **Université de Lille 2** Lille

✉ salahbeydjalil@gmail.com

🔗 <https://djo-project-portfolio.vercel.app/>

📍 France

☎ +33 6 11 27 91 53

Compétences Techniques

Data Platform & Processing

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

Data Storage & Databases

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

DataOps & CI/CD

Git · GitHub Actions · Azure DevOps · Docker · Kubernetes

Quality & Governance

Unity Catalog · Key Vault · Great Expectations

Monitoring & Observability

Grafana · Prometheus

Visualization & Analytics

Power BI · Streamlit · Metabase

Languages

Python · SQL · YAML · Bash

Cloud & Environment

Azure · GCP · Linux (WSL2)

Langues

Français

Anglais

Arabe

Espagnol

Centres d'intérêt

Sport

Tennis
Basket Ball
Football

Tech

Conférences Data
Veille IA