



Djalil Salah-bey

Data Engineer – Pipelines, Lakehouse & DataOps (Azure / GCP)

Data Engineer confirmé, spécialisé dans la conception et la mise en production de plateformes data modernes sur Azure Databricks et GCP BigQuery.

J'automatise et fiabilise les pipelines de données de bout en bout selon une approche DataOps.

Mes réalisations s'appuient sur PySpark, dbt, Airflow, Delta Lake et des environnements multi-cloud performants et gouvernés.

Passionné par la qualité, la collaboration et la valorisation métier de la donnée via Power BI et des modèles fiables et documentés.

Site web

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

Email

salahbeydjalil@gmail.com

Numéro de téléphone

+33 6 11 27 91 53

Langues

Français

Natif

Anglais

Courant

Passions

Sport

Programmation

Compétences Techniques

Langages : Python, PySpark, SQL, Shell

ETL/ELT : Airflow, dbt Core, Prefect, Talend

Bases de données :

PostgreSQL, MongoDB, InfluxDB, Firebase

Machine Learning : MLflow · scikit-learn · XGBoost ·

LightGBM · Tensorflow · SciPy

Visualisation : Power BI ·

Tableau · Looker Studio

CI/CD & Observabilité : Cloud

Build, Azure DevOps,

Prometheus, Grafana, Cloud

Monitoring / Logging, Docker

Data Platform: GCP, Azure, Databricks, Hadoop, Spark, Big Query.

Diplômes et formations

■ **Microsoft Azure DP-203** – Certification Data Engineer Associate (en cours)

■ **IA School** — Diplôme RNCP Niveau 7 Expert Data & Ingénierie de l'Intelligence Artificielle (2023-2025)

■ **IA School** — Bachelor Expert Data & Ingénierie de l'Intelligence Artificielle (2022-2023)

■ **Université de Lille** — Licence Banque, Finance (2019-2022)

Expériences professionnelles

PowerUp Technology

Data Engineer

Paris

De février 2025 à octobre 2025

- **Conception et orchestration de pipelines batch SFTP → ADLS → Delta Lake en PySpark**, avec contrôle de format, gestion des erreurs et traçabilité des fichiers de télémétrie clients.
 - **Développement de notebooks Databricks** pour les traitements métier : agrégations, enrichissements et calculs d'indicateurs de performance batterie.
 - **Orchestration Airflow (AKS)** : planification et exécution des jobs Databricks, dépendances, retries et alertes SLA.
 - **Tests automatiques PySpark** (formats, valeurs aberrantes, cohérence temporelle) garantissant la fiabilité en production.
 - **CI/CD Azure DevOps** : automatisation des validations, packaging et déploiement des notebooks, exécution des workflows et documentation versionnée.
 - **Observabilité Prometheus + Grafana** : intégration des logs Airflow/Databricks et tableaux de bord SLA/performances.
 - **Collaboration Data Science pour la** préparation et fiabilisation des datasets utilisés pour les modèles prédictifs batterie.
 - **Restitution métier sous Power BI** : suivi des performances, états de charge et alertes clients.
 - **Pratiques DataOps / MLOps** : CI/CD Databricks, versioning des artefacts et suivi des modèles en production.
 - **Intégration applicative** : KPI (PostgreSQL), configurations (MongoDB) et mesures techniques (InfluxDB).
- Stack** : Python · PySpark · Azure Databricks · Delta Lake · ADLS Gen2 · Airflow (AKS) · Azure DevOps · Prometheus · Grafana · Docker · PostgreSQL · MongoDB · InfluxDB · Power BI.

Koacher

Data & Analytics Engineer

Lyon

De décembre 2022 à décembre 2024

- **Conception et déploiement d'une plateforme analytique GCP** centralisant **Firebase**, **Stripe** et APIs sportives dans **BigQuery** (Bronze / Silver / Gold).
- **Ingestion automatisée** via **Cloud Composer (Airflow)** et **Dataflow (Beam)** : collecte quotidienne vers **GCS**, avec historisation et reprise sur incident.
- **Modélisation analytique dbt Core + BigQuery (Kimball, SCD1/2)** : dimensions utilisateurs, coaches, séances et faits sessions, paiements, performances.
- **Tests dbt et règles métier** : cohérence paiement-utilisateur, valeurs aberrantes et doublons pour fiabiliser les modèles.
- **Orchestration avancée Airflow** : DAGs multi-tâches, dépendances dynamiques, backfills et alerting SLA.
- **Documentation et lineage automatisés** via **dbt Docs** et métadonnées **BigQuery (INFORMATION_SCHEMA)**.
- **Observabilité native GCP** : supervision Composer/BigQuery avec **Cloud Monitoring / Logging**, dashboards coûts, latence et SLA.
- **CI/CD Cloud Build** : exécution automatique des tests dbt, génération de docs et déploiement multi-environnements.
- **Optimisation BigQuery** : partitioning, clustering et tuning SQL → -35 % coûts, -40 % temps d'exécution.
- **Collaboration Produit & Data Science** : modèle de données central et dashboards **Looker Studio / Power BI**.
- **Pratiques DataOps / Analytics Engineering** : industrialisation des pipelines dbt, CI/CD Cloud Build et observabilité continue.

Stack : Python · SQL · dbt Core · BigQuery · GCS · Airflow (Cloud Composer) · Firebase · Stripe API · Cloud Build · Cloud Monitoring · Cloud Logging · Docker · Looker Studio.

Muli By Easy Planet

Referent ETL & Flux

Paris

De janvier 2022 à juillet 2022

- Développement de **scripts Python** pour automatiser les chargements et contrôles de données.
 - Conception et maintenance de **flux ETL Talend** pour alimenter les bases **PostgreSQL**.
 - Intégration et planification des traitements dans **Airflow** pour un enchaînement fiable des jobs.
 - Contribution à la mise en **production et surveillance** des pipelines de données.
 - **Renforcement des bonnes pratiques DataOps** : tests, documentation et suivi de la qualité des flux.
- Stack** : Talend · Python · SQL · PostgreSQL · Airflow · Hadoop · Spark