



Djalil Salah-bey

Data & Analytics Engineer | En recherche de CDI

Data Engineer passionné par la programmation et l'automatisation des pipelines de données, j'aime concevoir et industrialiser chaque étape — ingestion, transformation, orchestration et supervision — pour garantir la fiabilité, la performance et la traçabilité des flux en production.

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

Passions

Sport

Programmation

Intelligence Artificielle

Compétences Techniques

Cloud & Plateformes :

Azure (ADLS, ADF, Databricks, Purview), GCP (BigQuery, Dataflow, Cloud Composer), Docker, AKS/GKE

Traitements :

Spark/PySpark, SQL, dbt (tests/docs), Delta Lake, Parquet, ClickHouse, Kafka

Orchestration & Qualité :

Airflow, Databricks Workflows, data contracts, Great Expectations

Observabilité & FinOps :

logs/métriques/traces, Prometheus/Grafana, coûts & sizing

Sécurité & RGPD :

RBAC/IAM, secrets, masquage/tokenisation PII, rétention & purge

CI/CD & IaC :

Github Actions, Azure DevOps, Terraform, tests unitaires & de régression

BI/Consommation :

Power BI, Looker Studio, APIs/exports contractuels

Langages :

Python (PySpark), SQL, Shell

Expériences professionnelles

● PowerUp Technology

Data Engineer

Paris

De novembre 2023 à octobre 2025

- **Conception et orchestration de pipelines batch** sous Airflow collectant les fichiers de mesures BESS déposés sur SFTP clients vers Azure Data Lake Gen2, avec gestion des dépendances, logs et reprises automatiques.
- **Mise en place d'un framework d'ingestion modulaire (Python + Airflow)** permettant d'onboarder de nouveaux clients via configuration dynamique des flux et métadonnées centralisées.
- **Développement et exécution** des traitements **PySpark** sur **Azure Databricks Workflows** pour l'intégration incrémentale, le calcul d'indicateurs et la mise à jour automatisée des jeux de données.
- **Calcul automatisé des KPI opérationnels** (performance, sécurité, prédition, usage batteries) issus des séries temporelles, utilisés pour le suivi client, la modélisation prédictive et l'intégration continue des modèles.
- **Conception d'un modèle hiérarchique et multi-dimensionnel** assurant le stockage incrémental et la restitution analytique des indicateurs batteries dans **PostgreSQL** via des schémas dédiés par client.
- **Développement d'une interface interne Python** pour la supervision et la gestion des métadonnées clients, adossée à **MongoDB** (métadonnées) et **InfluxDB** (séries temporelles).
- **Mise en place d'un système d'alertes métier automatisé** basé sur **InfluxDB** et **Grafana**, détectant en temps quasi réel les dérives thermiques et électriques des modules batteries et notifiant les clients par e-mail.
- **Industrialisation de la supervision technique** des pipelines Airflow et Databricks via **Grafana**, **Azure Monitor** et **Log Analytics**, garantissant la traçabilité, la disponibilité et la performance des workflows en production.
- **Optimisation Databricks via Delta Lake** (*time travel, schémas évolutifs, traitements incrémentaux*) et tuning des performances (*Z-Ordering, partitionnement, autoscaling*).
- **Mise en œuvre de la CI/CD** sous **Azure DevOps**, intégrant tests unitaires PySpark, validations métier et documentation automatique des workflows.

Stack : Python · PySpark · SQL · Airflow · Azure Databricks · ADLS Gen2 · Delta Lake · MLFlow · PostgreSQL · MongoDB · InfluxDB · Grafana · Azure DevOps · Docker · Git

● Koacher

Data Engineer

Lyon

De septembre 2022 à août 2023

- **Ingestion automatisée** des flux applicatifs (*Firebase, Stripe, Hubspot, APIs partenaires*) et capteurs sportifs via **Cloud Composer (Airflow)**, avec gestion des dépendances, contrôles de qualité et reprise automatique.
- **Structuration du Lakehouse** en zones **Bronze / Silver / Gold** (*GCS → BigQuery*) afin d'assurer la traçabilité, la fiabilité et la cohérence des données inter-sources.
- **Transformation et modélisation analytique** sous **dbt + BigQuery**, incluant la gestion des **SCD Type 2**, la création de **dimensions hiérarchiques** et de **tables de faits temporelles**, le développement de **modèles incrémentaux** avec tests automatiques et documentation générée dynamiquement.
- **Conception d'un modèle de données analytique unifié** combinant **métriques IoT** et **données applicatives**, offrant une vision 360° de la performance athlète et de l'engagement utilisateur.
- **Orchestration des traitements dbt** dans **Airflow (Cloud Composer)**, avec notifications, logs et dépendances multi-jobs.
- **Implémentation du CI/CD analytique** via **Github Actions** et **Cloud Build**, automatisant les tests, déploiements et documentation continue.
- **Optimisation BigQuery** : partitionnement temporel, clustering, tuning des coûts et suivi de la consommation par dataset.
- **Restitution analytique** dans **Looker Studio** et **Tableau**, avec tableaux de bord dédiés au pilotage sportif et à l'analyse produit.
- **Suivi de la qualité et de la conformité** : catalogage des datasets, documentation centralisée, conventions de nommage unifiées et supervision via **Cloud Monitoring**.

Stack : Python · SQL · dbt · Airflow (Cloud Composer) · BigQuery · GCS · Cloud Build · GitHub Actions · Looker Studio · Tableau · Docker · Kubernetes · Git

Diplômes et formations

- **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)