



Djalil Salah-bey

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

Data Engineer passionné par la construction de plateformes de données fiables et évolutives. J'aime concevoir des pipelines clairs, documentés et automatisés pour garantir la qualité et la disponibilité des données. Habitué aux environnements Azure Databricks et GCP BigQuery, j'applique au quotidien une approche DataOps basée sur la rigueur, la collaboration et l'amélioration continue.

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
Intelligence Artificielle

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)

Compétences Techniques

Langages : Python, R, SQL, Shell
Cloud & Plateformes : Azure (Databricks, ADLS, DevOps), GCP (BigQuery, GCS, Cloud Composer)
Orchestration & Modélisation : Airflow, dbt Core, Dataflow, Delta Lake
CI/CD & Observabilité : Cloud Build, Prometheus, Grafana, Cloud Monitoring
Bases & BI : PostgreSQL, MongoDB, InfluxDB, Power BI, Looker Studio
Big Data : Hadoop, Spark
IA : MLFlow, Sickit-Learn, Tensorflow, Keras, OpenCV, SciPy

Expériences professionnelles

- PowerUp Technology**
Data Engineer Freelance

Paris
De février 2025 à octobre 2025

 - Conception et déploiement d'une plateforme Lakehouse Azure (ADLS Gen2 + Databricks + Delta Lake) pour centraliser et historiser les flux IoT issus des systèmes de batteries connectées.
 - Développement de jobs PySpark dans Databricks (Delta Live Tables / Workflows) pour le nettoyage, la déduplication et l'agrégation des données brutes, avec gestion incrémentale et contrôle de schéma.
 - Modélisation analytique sous dbt + Databricks SQL, selon la méthode Kimball : création de dimensions hiérarchiques, faits temporels et SCD Type 2 (traçabilité complète des événements).
 - Mise en place d'un pipeline d'ingestion Azure Data Factory → ADLS → Databricks, orchestré par Airflow (AKS), avec gestion des dépendances, retries et monitoring d'exécution.
 - Implémentation de tests dbt et règles métier (Python + Great Expectations) pour garantir la qualité et la cohérence des mesures IoT (température, SOC, courant).
 - CI/CD GitHub Actions + Terraform (Azure Resources) : déploiement automatisé des notebooks Databricks, modèles dbt et ressources AKS / ADLS.
 - Supervision complète (Prometheus + Grafana) des jobs Databricks, pipelines Airflow et coûts Azure, avec alerting SLA et suivi de latence.
 - Développement de tables métiers agrégées exposées aux data scientists pour la modélisation thermique et la détection préventive d'anomalies

Stack : Azure Data Factory | ADLS Gen2 | Databricks (PySpark, SQL, Delta Lake, Workflows) | dbt | Airflow | ClickHouse | Terraform | GitHub Actions | Purview | Marquez | Prometheus | Grafana | Power BI.
- Koacher**
Data Engineer

Lyon
De décembre 2022 à décembre 2024

 - Conception et industrialisation d'une plateforme Lakehouse GCP (Bronze / Silver / Gold) centralisant les flux applicatifs, IoT et marketing pour fiabiliser et valoriser les données sportives.
 - Ingestion automatisée des flux applicatifs et IoT via Cloud Composer (Airflow) et Dataflow (Beam), intégrant gestion d'erreurs, reprise automatique et dépendances inter-systèmes.
 - Modélisation analytique sous dbt + BigQuery selon la méthode Kimball : création de dimensions hiérarchiques et tables de faits historisées (SCD Type 2).
 - Mise en œuvre de tests dbt et Great Expectations pour contrôler la qualité, la fraîcheur et la conformité des données sensibles (PII).
 - CI/CD et Infrastructure as Code (Terraform, GitHub Actions, Cloud Build) assurant le déploiement automatisé et la cohérence multi-environnements.
 - Supervision complète des jobs et pipelines (Airflow, Dataflow, BigQuery) via Cloud Monitoring + Grafana, avec alerting SLA et suivi de latence.
 - Publication de datasets analytiques dans BigQuery pour les équipes produit et marketing, alimentant automatiquement les dashboards Looker Studio et les KPIs d'engagement utilisateur.

Stack : Python | SQL | dbt Core | BigQuery | GCS | Airflow | Dataflow (Beam) | Terraform | GitHub Actions | Cloud Build | Cloud Monitoring | Grafana | Data Catalog | Looker Studio
- Muli By Easy Planet**
Développeur ETL

Paris
De janvier 2022 à juillet 2022

 - Développement de scripts Python pour automatiser les chargements, les contrôles d'intégrité et la préparation des jeux de données avant intégration.
 - Conception et maintenance de flux ETL Talend pour alimenter les bases PostgreSQL et les tables analytiques en respectant la logique incrémentale.
 - Orchestration des traitements sous Airflow, assurant la fiabilité, la reprise sur erreur et la planification quotidienne des jobs Talend.
 - Surveillance et optimisation des pipelines de données, avec mise en place de métriques de volumétrie et d'alerting automatisé.
 - Renforcement des pratiques DataOps avec gestion des référentiels, documentation, suivi qualité et normalisation des pipelines dans le Master Data Management MDM.
 - Création de tables d'exposition pour les équipes BI, facilitant la mise à jour automatisée des reportings de ventes et de performance produit.

Stack : Talend · Python · SQL · PostgreSQL · Airflow · Hadoop · Spark · Github Actions · Power BI · Tableau.