



Djalil Salah-bey

Data Engineer | Recherche CDI

Data Engineer spécialisé dans la conception et l'industrialisation de plateformes cloud Azure. Expérience sur des environnements Databricks, Snowflake et Airflow, avec une approche DataOps orientée performance, qualité et fiabilité des flux. Habitué à collaborer avec les équipes Data Science et Analytics pour transformer les données brutes en actifs exploitables à grande échelle.

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
Python · PySpark · SQL · Bash

Cloud & Plateformes
Microsoft Azure , Databricks , Google Cloud Platform

Bases de données
Oracle DB, SQL Server, MongoDB, InfluxDB, PostgreSQL

Orchestration & CI/CD
Airflow (AKS) · Azure Data Factory · GitHub Actions · Azure DevOps · Docker · Kubernetes · dbt Core

Visualisation & Moitoring
Power BI · Prometheus · Grafana · Terraform

Machine Learning & MLOps
MLflow · scikit-learn · XGBoost · LightGBM · Great Expectations

Diplômes et formations

- **Microsoft Azure DP –** Certification Data Engineer Associate (en cours)
- **IA School —** Diplôme RNCP Niveau 7 Expert Data & Ingénierie de l'Intelligence Artificielle (2022-2025)
- **IA School —** Bachelor Expert Data & Ingénierie de l'Intelligence Artificielle (2021-2022)
- **Université de Lille —** Licence Banque, Finance (2018-2022)

Soft Skills

- Curieux
- Proactif
- Rigoureux

Expériences professionnelles

PowerUp Technology

Freelance Data & AI Engineer

Paris

De février 2025 à octobre 2025

Mission de 8 mois : Amélioration et industrialisation de la **plateforme Lakehouse Azure Databricks** dédiée à la performance et à la maintenance prédictive des batteries industrielles (BESS).

- Refonte partielle du modèle **Bronze / Silver / Gold** sous **Delta Lake** pour fiabiliser les flux IoT (télémétrie, température, voltage).
- Automatisation de l'ingestion depuis **Azure IoT Hub** et APIs internes via **Databricks Auto Loader** et **Airflow (AKS)**.
- Rationalisation du stockage sur **ADLS Gen2** (partitionnement, Z-Order, compaction) afin d'améliorer la latence de lecture.
- Refactorisation des pipelines **PySpark / SQL** pour le nettoyage et l'enrichissement des données Silver.
- Création et documentation de tables **Gold** pour l'analyse de performance et la détection d'anomalies.
- Support aux équipes **Data Science** pour la préparation des jeux de features et le suivi des modèles sous **MLflow**.
- Maintenance et automatisation des workflows sous **Airflow (AKS)** et **Databricks Jobs**, avec mise en place d'alertes et gestion des SLA.
- **CI/CD GitHub Actions** : tests, packaging et déploiement automatisé des notebooks et configurations Databricks Repos.
- Contribution au déploiement infra Databricks via **Terraform** (clusters, permissions, monitoring).
- Suivi des métriques de jobs et coûts via **Grafana / Prometheus**, mise en place d'alertes Slack.
- Intégration de **Great Expectations** pour le contrôle qualité et la fraîcheur des datasets Silver / Gold.
- Travail en étroite collaboration avec les équipes **Data Science** (préparation, versioning datasets MLflow) et **Data Analytics** (tables Gold pour Power BI).

Résultats : Fiabilité pipelines > 99,5 % · Réduction de 30 % des temps de traitement · Standardisation des jeux de données Silver / Gold pour Data Science et Analytics.

Stack : Azure Databricks · Delta Lake · PySpark · Airflow (AKS) · MLflow · Terraform · GitHub Actions · Grafana · Prometheus · Great Expectations · ADLS Gen2 · Docker · Python · SQL.

Koacher

Alternant Data Engineer

Lyon

De décembre 2022 à décembre 2024

Conception et exploitation d'un **Customer Data Lakehouse Azure & Snowflake** destiné à centraliser et fiabiliser les flux applicatifs, marketing et produit selon une architecture **Bronze / Silver / Gold**.

- **Déploiement et administration d'Airflow sur AKS**, garantissant l'orchestration, le monitoring et la reprise automatique des pipelines à grande échelle.
- **Développement de pipelines d'ingestion** via **Azure Data Factory** et **Airflow**, intégrant des flux multi-sources (APIs, Firebase, Hubspot, Stripe) vers le Data Lake (Bronze).
- **Modélisation en flocon de neige sous dbt Core (SQL Jinja)** : gestion des **SCD Type 2**, création de dimensions et faits, application du modèle **Kimball** pour les couches Silver et Gold.
- **Optimisation Snowflake** (clustering, micro-partitioning, warehouse sizing) pour garantir des performances OLAP élevées et une maîtrise fine des coûts.
- **Mise en production CI/CD** (Azure DevOps + GitHub Actions) avec tests automatisés, déploiements contrôlés et documentation continue des modèles dbt et dashboards Power BI.
- **Conception de dashboards Power BI en Direct Query sur Snowflake**, reposant sur un modèle sémantique unifié et des mesures DAX standardisées.
- **Mise en place d'un catalogue KPI** (rétention, churn, LTV, engagement) et d'un **système d'alerting qualité** (ADF + Slack + Power BI).
- **Supervision technique** via **Grafana et Prometheus** assurant le suivi des SLA des jobs **Airflow & Azure Data Factory**.
- **Lineage et gouvernance assurés via Marquez et Purview**, garantissant la traçabilité et la conformité des datasets critiques.

Résultats : -35 % de coûts Snowflake, -50 % d'incidents d'ingestion, dashboards actualisés en < 15 min.

Stack : Azure Data Factory · Airflow (AKS) · Snowflake · dbt Core · Power BI · Azure DevOps · Purview · Marquez · Python · SQL · Grafana · Prometheus · Git · Docker · Kubernetes.