



Djalil Salah-bey

Data Engineer | En recherche de CDI

Data Engineer passionné par la programmation et l'automatisation, j'aime concevoir, déployer et fiabiliser des pipelines de données assurant la disponibilité, la traçabilité et la qualité des flux en environnement industriel.

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

Cloud & Data Platform

- Azure, GCP, Databricks

Orchestration & Scheduling

- Airflow (AKS, GKE), Databricks Workflows

Modélisation & Transformation

- dbt, SQL, PySpark, Pandas

Monitoring & Observabilité

- Grafana, Prometheus, OpenLineage, Marquez

Databases & BI

- PostgreSQL, MongoDB, DuckDB, ClickHouse
- Looker Studio, Power BI

Développement & Conteneurisation

- Python, SQL, Bash
- Docker, Kubernetes

Diplômes et formations

- IA School —Paris Licence 3 & Mastère Expert en Data Engineering & Intelligence Artificielle (RNCP Niv.7, 2022 – Avril 2025)
- IAE de Lille— Master Système d'information (2021-2022)
- Université de Lille — Licence Banque, Finance (2018-2021)

Expériences professionnelles

PowerUp Technology

Freelance Data Engineer - 9 mois

Paris

De février 2025 à octobre 2025

Conception, déploiement et exploitation d'une **plateforme Data Lakehouse sur Azure** pour automatiser la collecte, le traitement et l'exposition de données issues de systèmes BESS (Battery Energy Storage Systems).

- Conception et structuration du Data Lake sur Azure Data Lake Storage Gen2**, organisé en zones Bronze / Silver / Gold, avec stockage Delta / Parquet et stratégies de rétention adaptées aux besoins analytiques.
- Orchestration de pipelines Airflow déployés sur Azure Kubernetes Service (AKS)**, assurant la planification, la supervision des SLA et le déclenchement des Databricks Workflows dans le cadre de déploiements continus via Azure DevOps Pipelines.
- Développement et optimisation de traitements distribués sous Azure Databricks (Python / PySpark)**, intégrés aux DAGs Airflow pour le nettoyage, la consolidation et la modélisation des données en production.
- Industrialisation de modèles analytiques dbt Core exécutés sous Airflow**, avec tests automatisés, documentation centralisée et génération d'une couche sémantique unifiée pour les analyses BI et Data Science.
- Implémentation d'un dispositif de supervision et de fiabilité des flux sous Prometheus (connecté à Airflow Metrics)**, assurant la visibilité en temps réel des pipelines et la stabilité des traitements en production.
- Développement de librairies Python internes** pour mutualiser les fonctions d'ingestion, de validation et de logging dans les pipelines, améliorant la maintenabilité et réduisant la duplication de code.
- Intégration et automatisation de modules de Machine Learning et d'algorithmes analytiques dans les pipelines Databricks**, en collaboration avec les Data Scientists, pour exécuter les workflows d'inférence et calculer des indicateurs de performance, d'usage, de sécurité et de vieillissement.
- Contribution à la documentation technique, au suivi du data lineage et à la gestion du code sous Azure DevOps Repos**, avec revues de code et diffusion des bonnes pratiques Data Engineering.

Stack technique : Python · SQL · PySpark · Azure Databricks · ADLS Gen2 · Airflow (AKS) · dbt · Azure DevOps (Repos & Pipelines) · Prometheus · Docker · PostgreSQL · ClickHouse · MongoDB · Unity Catalog · Azure Wiki.

Koacher

Alternant Data Engineer - 2 ans

Lyon

D'octobre 2022 à octobre 2024

Développement, évolution et exploitation de la **plateforme analytique de l'application mobile Koacher** basée sur GCP pour transformer les données d'usage en indicateurs produit et sportifs.

- Développement et évolution continue de pipelines ELT en Python sous Airflow**, avec écriture des tâches, gestion des dépendances, appels aux APIs applicatives, parsing de payloads JSON et chargement automatisé des données normalisées dans BigQuery.
- Conception et développement de modèles analytiques BigQuery sous dbt Core**, avec structuration des couches **staging / intermédiaire / marts (datamarts analytiques)**, écriture de requêtes SQL complexes, création de macros et de tests dbt, logiques d'historisation et définition d'une **couche sémantique unifiée** utilisée par les équipes Produit, BI et Data.
- Structuration et optimisation continue du stockage analytique dans BigQuery**, en ajustant via le code le partitionnement, le clustering et les politiques de rétention afin d'améliorer les performances de requêtes et de maîtriser les coûts cloud.
- Développement de librairies internes Python** (ingestion, validation, logging, monitoring), organisées en modules réutilisables et intégrés dans les DAGs Airflow pour standardiser les patterns de traitement et réduire la duplication de code.
- Préparation et exposition de jeux de données features et labels dans BigQuery** pour alimenter les modèles de Machine Learning développés par les Data Scientists, et **intégration** des résultats de scoring dans les modèles dbt consommés par les équipes Produit.
- Mise en place et maintenance de pipelines CI/CD GitLab** pour les projets dbt et Airflow, avec écriture des pipelines (YAML), intégration de tests unitaires Python, tests dbt, linting et déploiement automatisé sur les environnements Dev et Prod.
- Instrumentation et observabilité des pipelines via Grafana**, avec exposition de métriques techniques (durée, volumétrie, fraîcheur, erreurs) et configuration d'alertes sur les SLA pour sécuriser l'exécution en production.
- Gestion du code et documentation technique sous GitLab Repos**, avec revues de code, utilisation de dbt Docs et formalisation des bonnes pratiques de développement DataOps (structure des projets, conventions de nommage, qualité du code).

Stack technique : Python · SQL · dbt Core · Airflow (Cloud Composer) · BigQuery · GCP (Cloud Storage) · GitLab CI/CD · Grafana · Docker · PostgreSQL