



Djalil Salah-bey

Data & Analytics Engineer | En recherche de CDI

Data Engineer expert en pipelines ETL/ELT automatisés, en modélisation analytique et en DataOps de production. 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 (Databricks, ADLS Gen2), GCP (BigQuery, GCS)

Orchestration & Scheduling

- Airflow (AKS, GKE), Databricks Jobs/Workflows

Modélisation & Transformation

- dbt (modèles métiers, tests, documentation)
- SQL, PySpark, Pandas

DataOps & Qualité

- GitHub Actions (CI/CD), Terraform (IaC)
- dbt tests, Pytest, Great Expectations

Monitoring & Observabilité

- Grafana, Prometheus, OpenLineage, Marquez

Databases & BI

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

Développement & Conteneurisation

- Python, SQL, Bash
- Docker, Kubernetes (minikube, AKS)

Diplômes et formations

- IA School — Paris Licence 3 & Mastère Expert en Data Engineering & Intelligence Artificielle (RNCP Niv.7, 2022 – 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 et structuration du Data Lake sur Azure Data Lake Storage Gen2**, organisé en zones Bronze / Silver / Gold, avec gestion Delta / Parquet, partitionnement logique et politiques de rétention garantissant performance et traçabilité.
- Développement et orchestration des pipelines Airflow déployés sur Azure Kubernetes Service (AKS)**, assurant la planification, la fiabilité et la continuité des traitements avec tests et déploiements automatisés via Azure DevOps Pipelines.
- Développement et optimisation de traitements distribués sous Azure Databricks (PySpark, Delta Lake)**, orchestrés par Databricks Jobs et intégrés aux workflows Airflow, assurant performance, fiabilité et scalabilité des transformations en production.
- Industrialisation des modèles analytiques dbt intégrés au pipeline Airflow**, avec exécution planifiée, tests automatisés et documentation centralisée, garantissant une couche sémantique stable et cohérente exposée aux équipes BI.
- Supervision et suivi de la qualité des pipelines via Grafana et Airflow Metrics**, permettant le contrôle des SLA, la traçabilité des traitements et la surveillance continue de la fraîcheur des données.
- Optimisation du stockage et des performances analytiques sur Delta Lake**, avec mise en place d'une stratégie de partitionnement, d'indexation et de clustering, garantissant rapidité des requêtes et maîtrise des coûts cloud.
- Développement de bibliothèques Python internes** pour standardiser l'ingestion, la validation et le logging des traitements Data, renforçant la maintenabilité et la cohérence du code en production.
- Conception et intégration d'algorithmes analytiques avancés (Python / PySpark)** pour le calcul d'indicateurs de performance, de fiabilité et de vieillissement, enrichissant la couche analytique et les dashboards BI.

Stack technique :

Azure Databricks · ADLS Gen2 · Airflow (AKS) · PySpark · Delta Lake · dbt · Azure DevOps (Repos & Pipelines) · Grafana · Docker · PostgreSQL · ClickHouse · MongoDB · Unity Catalog

Koacher

Alternant Data Engineer – 2 ans

Lyon

D'octobre 2022 à octobre 2024

- Conception, évolution et maintenance des modèles de données sous dbt**, avec mise en place et enrichissement d'une **couche sémantique unifiée** exposée aux équipes BI et métiers.
- Construction et industrialisation progressive de pipelines ELT sous Airflow**, automatisant la collecte, la transformation et le chargement de données ERP & CRM dans le respect des SLA quotidiens.
- Rédaction des spécifications techniques et conception des modèles de données** garantissant la qualité, la traçabilité et la cohérence des jeux de données livrés.
- Généralisation des tests automatisés (dbt tests, Pytest)** et mise en place d'une **chaîne CI/CD GitLab** (lint, tests, déploiement) pour renforcer la fiabilité, la maintenabilité et la revue de code des pipelines.
- Optimisation continue des performances** (requêtes SQL, coûts BigQuery, parallélisation Airflow) avec réduction mesurable des temps de traitement et amélioration de la disponibilité des datasets.
- Implémentation de dashboards de suivi Airflow / Grafana / Prometheus** pour monitorer fraîcheur, volumétrie, erreurs et SLA, avec alerting proactif et baisse durable des incidents en production.
- Documentation technique et gouvernance des données** : normalisation des conventions (naming, lineage, accès), mise à jour des dictionnaires et runbooks, maintien continu de la couche sémantique dbt.
- Accompagnement technique des équipes BI et Data Analysts** sur l'usage des modèles dbt / SQL et l'exploitation des datasets BigQuery en self-service.

Stack Technique:

GCP (BigQuery, Cloud Composer, Cloud Storage) · Airflow · dbt · SQL · Python · GitLab CI/CD · Grafana · Prometheus · Talend · PostgreSQL · MongoDB

Projet personnel -Depuis Octobre 2025

Pipeline DataOps entièrement automatisé sous Airflow, avec **scripts Python d'ingestion** des APIs météo / pollution / trafic, traitements PySpark et modélisation analytique sous dbt (backend DuckDB), jusqu'à l'actualisation des dashboards Grafana. Déploiement en environnement conteneurisé proche production (Docker / Kubernetes), avec une chaîne de **CI** GitHub Actions pour les tests (dbt, Pytest, lint) et de **CD** pour le build et le déploiement automatisé des conteneurs.

Stack : Airflow · Python · PySpark · dbt · DuckDB · Grafana · Marquez · GitHub Actions · Docker · Kubernetes

Code : github.com/IADJALILProject/iot-smartcity-data-platform