



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, déployer et industrialiser chaque étape pour garantir la fiabilité, la performance, la traçabilité et la disponibilité 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

Compétences Techniques

Cloud & Data Platform

- Azure (Databricks, ADLS Gen2), GCP (BigQuery, GCS)

Orchestration & Scheduling

- Airflow (AKS, Cloud Composer), 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 — Diplôme RNCP Niveau 7 Data Science & expert IA (2022-2024)
- IAE de Lille — Master Système d'information & management de la donnée (2021-2022)
- Université de Lille — Licence Banque, Finance (2018-2021)

Expériences professionnelles

PowerUp Technology

Freelance Data Engineer

Paris

De février 2025 à octobre 2025

Chargé de la mise en production d'une **plateforme IoT multi-clients** sur **Azure Databricks Lakehouse** dédiée à l'analyse de performance et de vieillissement des batteries BESS.

- Ingestion & orchestration** : automatisation des flux (SFTP et bases clients) via **Airflow AKS** et **Databricks Workflows**, avec gestion des dépendances, reprises sur erreur et exécution planifiée sans intervention manuelle.
- Traitement & normalisation** : développement de jobs **PySpark** pour le nettoyage, la validation et la structuration des signaux capteurs avec contrôles intégrés de complétude et cohérence.
- Modélisation analytique** : création de modèles **dbt Databricks** pour le calcul des indicateurs de performance, sécurité et vieillissement avec stratégie incrémentale, tests et documentation automatisée.
- Exposition & actualisation** : intégration automatisée des tables analytiques **Gold** dans les **dashboards clients** garantissant la mise à jour quotidienne et la cohérence des indicateurs métier.
- Supervision & DataOps** : suivi des jobs et indicateurs de qualité via **Grafana / Prometheus** et automatisation des déploiements et tests sous **Azure DevOps**, assurant traçabilité, fiabilité et continuité du run.

Stack : Azure Databricks · Airflow (AKS) · ADLS Gen2 · PySpark · dbt · SQL · Delta Lake · PostgreSQL · MongoDB · InfluxDB · Grafana · Prometheus · GitHub Actions · Docker

Koacher

Data Engineer

Lyon

De septembre 2022 à novembre 2024

- Ingestion & orchestration** : automatisation des flux de données (**application mobile Firebase**, **paiements Stripe**) vers **Google Cloud Storage** et **BigQuery**, orchestrée via **Cloud Composer (Airflow)** avec planification quotidienne, gestion des dépendances et retries.
- Traitement & préparation** : transformation et normalisation des événements via **BigQuery SQL** et **Dataflow**, avec règles de nettoyage et de standardisation des schémas.
- Modélisation analytique** : conception de modèles **dbt sur BigQuery** pour structurer les faits et dimensions avec modèles incrémentaux, SCD Type 2 et tests intégrés.
- Exposition & actualisation** : alimentation automatique des dashboards **Looker Studio** à partir des tables BigQuery garantissant une mise à jour quotidienne pour les équipes coaching et business.
- Data Quality & monitoring** : mise en place de tests **dbt** (fraîcheur, unicité, relations) déclenchés par **Cloud Composer** et suivi des exécutions via les logs GCP, assurant stabilité et fiabilité des pipelines en production.

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

Projet perso – Septembre 2025

- Développement d'un pipeline IoT complet orchestré par Airflow**, collectant et historisant les flux publics OpenWeatherMap, OpenAQ et Transport Data Gouv pour la **supervision de la qualité de l'air et du trafic sur Paris**.
- Ingestion automatisée** en modes batch et micro-batch, avec transformations PySpark en zone Silver : nettoyage, normalisation, partitionnement et historisation optimisée.
- Modélisation analytique dbt** : création d'un schéma en étoile consolidé (météo-pollution-traffic), stratégie incrémentale et tests de qualité intégrés.
- CI/CD et tests automatisés** : Makefile, GitHub Actions, dbt/Pytest et conteneurisation Docker/Kubernetes assurant qualité, portabilité et stabilité en production.
- Supervision et observabilité temps réel** : dashboards Grafana pour SLA et fraîcheur des données, traçabilité complète via OpenLineage / Marquez.

Stack : Airflow · Python · PySpark · dbt · DuckDB · Parquet · Grafana · OpenLineage · Marquez · GitHub Actions · Makefile · Docker · Kubernetes · Pytest

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