



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 Expert Data & Ingénierie de l'Intelligence Artificielle (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 septembre 2025

- Conception et mise en production d'une plateforme sur Azure Databricks** centralisant les flux IoT multi-clients pour le suivi de performance et de vieillissement des batteries industrielles.
- Ingestion et orchestration automatisées** (Airflow AKS, Databricks Workflows) : pipelines journaliers sans intervention manuelle, gestion des dépendances et reprise sur erreur.
- Nettoyage, normalisation et enrichissement PySpark** des signaux capteurs (température, tension, SOC, SOH) avec contrôles de complétude et de cohérence des données.
- Modélisation analytique dbt** : calcul des indicateurs de performance, vieillissement, sécurité et usage, intégrant stratégie incrémentale, tests automatiques et CI/CD via GitHub Actions.
- Exposition automatisée des tables analytiques** vers les dashboards clients via API, assurant actualisation continue et cohérence des indicateurs batterie.
- Optimisation des performances et des coûts Databricks** : autoscaling des clusters, paramétrage du cache et suivi des coûts Grafana pour améliorer les SLA et réduire la latence.
- Infrastructure as Code et sécurité cloud** : déploiement automatisé des environnements Data (Terraform, Azure DevOps) et gestion des permissions RBAC garantissant isolation et conformité des espaces client.

Stack : Python · PySpark · SQL · dbt · Azure Databricks · Azure Kubernetes Services · Airflow · Azure Data Lake Gen2 · Delta Lake · PostgreSQL · MongoDB · InfluxDB · Grafana · Prometheus · Azure DevOps · Docker · Git

Koacher

Data Engineer

Lyon

De septembre 2022 à novembre 2024

- Conception et déploiement d'un Data Lakehouse GCP** (GCS → BigQuery → dbt) pour centraliser et historiser les données issues des capteurs sportifs et des applications mobiles.
- Optimisation du stockage et des performances BigQuery** : partitionnement, clustering et stratégies de coût garantissant scalabilité et requêtes efficaces.
- Ingestion et orchestration automatisées** sous Cloud Composer (Airflow) et Cloud Functions, assurant des flux quotidiens fiables, traçables et sans intervention manuelle.
- Framework ETL/ELT modulaire** : gestion des dépendances, logs et reprises sur erreur pour fiabiliser et industrialiser l'ensemble des pipelines.
- Transformation et modélisation analytique dbt** : nettoyage, normalisation et schéma en étoile consolidant utilisateurs, capteurs et paiements avec versioning SCD Type 2 et contrôles qualité intégrés.
- Gouvernance et DataOps** : documentation automatisée (dbt Docs), tests et CI/CD via GitHub Actions garantissant qualité, traçabilité et cohérence du modèle de données.
- Supervision et observabilité** : suivi de la qualité des pipelines et des coûts via Stackdriver, alerting sur erreurs et reporting automatisé de la performance GCP.
- Restitution BI** : dashboards Looker Studio pour le suivi des performances utilisateurs, la monétisation et la segmentation des abonnements.

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