



Djalil Salah-bey

Data & Analytics Engineer | En recherche de CDI

Data Engineer spécialisé en plateformes cloud-native (Azure & GCP), 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

Diplômes et formations

- **IA School** — Diplôme RNCP Niveau 7 Expert en science des données & développement de 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)

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)

Expériences professionnelles

- **PowerUp Technology -EnergyTech** Freelance Data Engineer Paris
De février 2025 à octobre 2025
 - **Conception et mise en production d'un pipeline de bout en bout** pour l'ingestion quotidienne des données clients par SFTP orchestré par Airflow sur AKS avec traçabilité complète des traitements et logs d'exécution.
 - **Structuration du Data Lake ADLS** selon une architecture en zones (**Bronze, Silver, Gold**) avec règles de nommage, schéma logique par client et partitionnement pour les **traitements distribués Databricks**.
 - **Développement de jobs PySpark sur Azure Databricks** pour le nettoyage, la validation et la standardisation des flux de données (*courant, T°C, SOC, SOH, cycles*) en gérant schémas variables et données manquantes.
 - **Industrialisation du calcul des indicateurs batteries** : performance, sécurité, vieillissement et usage, exposés sous forme de **tables analytiques** consommables par les dashboards clients.
 - **Alimentation automatique des dashboards clients** : intégration des jeux de données dans les modèles visualisés (heatmaps température vs état de charge, courbes SOH, tuiles KPI) sans intervention manuelle.
 - **Mise en place de contrôles de qualité et d'alerting automatisé** (complétude, cohérence physique, anomalies) pour fiabiliser les analyses Produit & Client.
 - **Optimisation des traitements Databricks** (partitionnement, parallélisation, tuning des ressources) pour réduire le temps de calcul et les coûts cloud.
 - **Maintenance corrective et évolutive** : analyse d'incidents, mise en place de mécanismes de reprise (**retry, re-run, backfill**) et amélioration continue des traitements en production.
 - **Chaîne CI/CD sous Azure DevOps** : intégration des tests, déploiements Databricks et scripts Airflow automatisés via pipelines YAML.
 - **Collaboration étroite avec les équipes Produit, Data Science et Front** pour traduire les besoins (suivi de santé, détection de risques, performance) en modèles de données exploitables.

Stack : Azure (ADLS, Databricks, AKS, DevOps) · Airflow · PySpark · Python · SQL · PostgreSQL · Git · CI/CD · MongoDB · influxDB
- **Koacher -Sportech** Data Engineer Lyon
De novembre 2022 à novembre 2024
 - **Conception et industrialisation des modèles de données métiers dans dbt**, avec couche sémantique documentée et exposée aux équipes BI / Data.
 - **Mise en production et orchestration de pipelines ELT sous Airflow** alimentant **BigQuery** depuis des sources hétérogènes (*APIs, SFTP, bases SQL, fichiers GCS*) avec planification quotidienne et gestion des dépendances.
 - **Renforcement de la fiabilité des traitements en production** : tests dbt / Pytest, revues de code GitLab et **CI/CD** automatisant les déploiements entre environnements et les contrôles de qualité avant passage en prod.
 - **Optimisation des performances et des coûts BigQuery** (*partitionnement, clustering, tuning SQL*) avec réduction mesurable des temps de requête et de la facturation cloud.
 - **Mise en place de dashboards de supervision** (*Airflow, Stackdriver, Grafana*) pour le suivi des **SLA, de la fraîcheur, de la volumétrie** et du taux de succès, avec **alerting proactive** en cas d'incident.
 - **Maintenance corrective et évolutive** des pipelines : analyse d'incidents, mise en place de mécanismes de reprise (*retry, backfill*) et adaptation continue aux nouveaux besoins métiers.
 - **Intégration de nouvelles sources de données** : analyse des besoins, cartographie des flux, modélisation cible puis mise en production dans la couche sémantique **BigQuery/dbt**.
 - **Contribution à la gouvernance des données** : conventions de nommage, gestion des droits d'accès BigQuery, dictionnaire de données, runbooks d'exploitation et accompagnement des équipes BI / Data.

Stack : GCP (BigQuery, Cloud Storage, Cloud Composer) · Airflow · dbt · SQL · Firebase · Stripe · Python · GitLab CI/CD · Grafana · Stackdriver

Projet Sélectionné

- Stack :** Airflow · Python · PySpark · dbt (DuckDB) · Parquet · Grafana · Marquez · GitHub Actions · Docker / Kubernetes
- **Orchestration Airflow** collectant automatiquement les flux météo, pollution et trafic routier (*APIs publiques*).
 - **Ingestion Python (Bronze) et transformation PySpark (Silver)** stockées en Parquet optimisé.
 - **Modélisation dbt sous DuckDB** avec incrémentaux, tests automatiques et documentation intégrée.
 - **CI/CD complet** via Makefile & GitHub Actions : tests, build et déploiement conteneurisés (Docker + Kubernetes).
 - **Supervision Grafana et lineage Marquez** assurant le suivi des SLA, de la fraîcheur et des alertes pollution.
 - **Pipeline opérationnel 24/7**, totalement autonome et monitoré.

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