



# 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 -Scale-up EnergyTech (batteries BESS)

Data Engineer -9mois

Paris  
De février 2025 à octobre 2025

- Conception et mise en production** d'une architecture Data de bout en bout pour la supervision des batteries clients : ingestion SFTP orchestrée par **Airflow sur AKS**, historisation dans **Azure Data Lake (Bronze / Silver / Gold)** et transformations distribuées sous **Databricks**.
- Développement et industrialisation de traitements PySpark** pour le nettoyage, la validation et la standardisation des télémétries batteries (*courant, tension, température, SOC, SOH, cycles, RUL, HVAC*) avec gestion automatique des schémas variables et valeurs manquantes.
- Modélisation analytique et calcul d'indicateurs physiques** : performance, sécurité, vieillissement, usage et rendement énergétique, exposés dans **PostgreSQL** via une modélisation multidimensionnelle optimisée pour la visualisation et le reporting.
- Alimentation automatisée des dashboards clients** : intégration quotidienne des données agrégées vers les modèles front (*heatmaps, courbes SOH, KPI*) et synchronisation des jeux de données analytiques via Airflow et CI/CD Azure DevOps.
- Mise à disposition et versioning des jeux de données pour la Data Science** : création de jeux consolidés et horodatés servant à l'entraînement et au recalcul périodique des modèles prédictifs (*vieillessement, Remaining Useful Life, détection d'anomalies*).
- Support MLOps** : automatisation des workflows de scoring batch sous Airflow, intégration continue des modèles validés (Data Science) dans la chaîne de production, suivi des performances et recalibrage automatisé.
- Mise en place de contrôles de qualité et monitoring** : règles de complétude, cohérence physique, supervision Grafana / Azure Monitor, alerting proactif sur les SLA et les anomalies de données.
- Maintenance corrective et évolutive** : backfill automatisé, reprise sur incident, refactoring des jobs et amélioration continue de la scalabilité Databricks.
- Chaîne CI/CD sous Azure DevOps** : pipelines YAML automatisant les tests unitaires, le déploiement des notebooks Databricks et la publication des DAGs Airflow.
- Collaboration étroite avec les équipes Produit, Data Science et Software** pour traduire les besoins métiers en modèles de données exploitables et assurer la cohérence entre couche Data, ML et application client.

**Stack** : Azure (ADLS Gen2, Databricks, AKS, DevOps, Monitor) · Airflow · PySpark · Python (pandas, MLflow) · SQL · PostgreSQL · Grafana · Azure DevOps (CI/CD) · DataOps & MLOps.

#### Koacher - Scale-up SaaS (coaching & formation digitale)

Data Engineer -2ans

Lyon  
De novembre 2022 à novembre 2024

- Conception et mise en place de la plateforme Data Produit** sur **GCP**, centralisant les événements applicatifs (**Firestore**), les transactions **Stripe**, le CRM et les données back-end dans **BigQuery**.
- Construction de pipelines d'ingestion et de consolidation** (APIs, webhooks, fichiers batch) orchestrés sous **Airflow** avec historisation des événements utilisateurs (sessions, inscriptions, achats, complétion de parcours).
- Modélisation analytique sous dbt** : création de modèles métier et de data marts (utilisateurs, sessions, parcours, revenus, churn, engagement) avec tests automatiques et documentation intégrée.
- Mise en place d'une base analytique unifiée** exposée aux équipes Produit, Marketing et Sales via BigQuery et outils de BI, avec segmentation fine des comportements utilisateurs (cohorte, funnel, rétention, LTV).
- Préparation et versioning de datasets pour l'IA** : construction de jeux de données consolidés pour les Data Scientists (features comportementales, historiques d'usage, historique de paiement) utilisés pour les modèles de churn, de scoring lead et de recommandation de contenu.
- Support MLOps** : automatisation des workflows de scoring dans Airflow (churn, recommandation de contenu, appétence offre), actualisation régulière des scores en base et exposition des résultats vers l'app mobile.
- Fiabilisation en production** : mise en place de contrôles de qualité (fraîcheur, complétude, cohérence), mécanismes de reprise (**retry, backfill**) et suivi des SLA des principaux pipelines.
- Chaîne CI/CD sous GitLab** : validation de code, exécution des tests dbt et déploiement automatisé des modèles et DAGs vers les environnements dev → prod.
- Collaboration étroite avec les équipes Produit, Growth et Data Science** pour traduire les cas d'usage (activation, rétention, recommandation, monétisation) en modèles de données et workflows décisionnels.

**Stack** : GCP (BigQuery, Cloud Storage, Cloud Composer) · Firestore · Stripe · Airflow · dbt · SQL avancé · Python (pandas) · GitLab CI/CD · Stackdriver / Grafana · MLOps (workflows de scoring ML orchestrés sous Airflow)

## Projet Sélectionné

**Stack** : Airflow · Python · PySpark · dbt (DuckDB) · Parquet · Grafana · Marquez · GitHub Actions · Docker / Kubernetes

- **Orchestration Airflow** collectant tous les jours les flux météo, pollution et trafic routier sur Paris (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](https://github.com/IADJALILProject/iot-smartcity-data-platform)