



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

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 -Scale-up EnergyTech (batteries BESS)**

Data Engineer

Paris

D'octobre 2023 à décembre 2025

 - Conception et mise en production** d'une architecture data de bout en bout pour la supervision des batteries clients : ingestion quotidienne via **SFTP** orchestrée par **Airflow sur AKS**, historisation dans **Azure Data Lake** (zones Bronze/Silver/Gold) et transformations distribuées sous **Databricks**.
 - Développement et industrialisation de traitements PySpark** pour le nettoyage, la validation et la standardisation des télémetries batteries (courant, tension, température, SOC, SOH, cycles, RUL, HVAC), avec gestion automatique des schémas variables et des 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 et synchronisation des modèles front (heatmaps, courbes SOH, KPI) via Airflow et la chaîne CI/CD Azure DevOps, sans intervention manuelle.
 - 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 par les équipes Data Science dans la chaîne de production, suivi des performances et recalibrage automatisé.
 - Mise en place de contrôles de qualité et de monitoring** : règles de complétude et de cohérence physique, supervision sous Grafana / Azure Monitor et 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 · ClickHouse · MongoDB · influxDB · PostgreSQL · Grafana · Azure DevOps (CI/CD)
- **Koacher - Scale-up SaaS (coaching & formation digitale)**

Data Engineer -2ans

Lyon

De septembre 2022 à août 2023

 - Conception et mise en place d'une plateforme Data Produit** sur **GCP**, centralisant les événements applicatifs de **Firestore**, les transactions **Stripe**, le CRM et les données back-end dans **BigQuery**.
 - Développement de pipelines d'ingestion et de consolidation** (APIs, webhooks) orchestrés sous **Airflow** assurant l'historisation des événements utilisateurs : sessions, inscriptions, achats et complétion de parcours.
 - Modélisation analytique DBT** pour la création de modèles métier et de data marts utilisateurs, sessions, parcours, revenus, churn et engagement avec tests automatiques et documentation intégrée pour fiabiliser KPI Produit & Growth.
 - Construction d'une base analytique unifiée** exposée aux équipes Produit, Marketing et Sales via BigQuery et outils de BI, permettant la segmentation comportementale (cohortes, funnels, rétention, LTV).
 - Préparation et versioning des datasets IA** : création de jeux de données consolidés pour les équipes Data Science (features comportementales, historiques d'usage, paiements), alimentant les modèles de churn, de scoring et de recommandation de contenu.
 - Support MLOps** : automatisation des workflows de scoring batch dans Airflow (churn, recommandation, appétence offre), actualisation périodique des scores et exposition directe des résultats dans l'application.
 - Fiabilisation en production** : mise en place de contrôles de qualité sur la fraîcheur, la complétude et la 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 de développement et de production.
 - Collaboration étroite avec les équipes Produit, Growth et Data Science** pour transformer les cas d'usage (activation, rétention, recommandation, monétisation) en modèles de données et workflows décisionnels exploitables.

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)

Projets sélectionnés

- **PowerUp Technology – Pipeline de supervision batteries (Projet professionnel)**

Pipeline complet de supervision batteries – ingestion **SFTP** → **ADLS** → **ClickHouse**, exposition via **API Streamlit** et visualisation en temps réel sous **Grafana**, orchestrée par **Airflow**.

Génération automatisée des heatmaps température / cycle (charge – décharge – repos) et mise à jour continue des dashboards clients.

Code : github.com/IADJALILProject/heatmap_migration

- **IoT Smart City – DataOps Platform (Projet personnel)**

Pipeline IoT complet orchestré par **Airflow**, collectant les flux météo, pollution et trafic routier (APIs OpenWeatherMap, OpenAQ, Transport Data Gouv).

Ingestion batch & micro-batch, transformations **PySpark**, modélisation analytique **dbt** / **DuckDB**, monitoring **Grafana** / **Marquez** et CI/CD **GitHub Actions** / **Docker** / **Kubernetes**.

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