



# 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 et industrialiser chaque étape pour garantir la fiabilité, la performance et la traçabilité 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

### Passions

Sport  
Programmation  
Intelligence Artificielle

### Compétences Techniques

- Cloud & Plateformes : Azure (ADLS, ADF, Databricks, Purview), GCP (BigQuery, Dataflow, Cloud Composer), Docker, AKS/GKE
- Traitements : Spark/PySpark, SQL, dbt (tests/docs), Delta Lake, Parquet, ClickHouse, Kafka
- Orchestration & Qualité : Airflow, Databricks Workflows, data contracts, Great Expectations
- Observabilité & FinOps : logs/métriques/traces, Prometheus/Grafana, coûts & sizing
- Sécurité & RGPD : RBAC/IAM, secrets, masquage/tokenisation PII, rétention & purge
- CI/CD & IaC : GitHub Actions, Azure DevOps, Terraform, tests unitaires & de régression
- BI/Consommation : Power BI, Looker Studio, APIs/exports contractuels
- Langages : Python (PySpark), SQL, Shell

### Expériences professionnelles

- PowerUp Technology

Freelance Data Engineer - 9 mois

Paris

De février 2025 à octobre 2025
- Conception et orchestration de pipelines batch Airflow déployés sur AKS pour collecter les fichiers de mesures BESS déposés sur les SFTP clients vers Azure Data Lake Gen2, avec gestion des dépendances, journalisation et reprise automatique des workflows.
  - Conception d'un framework d'ingestion modulaire en Python et Airflow, appuyé sur MongoDB et InfluxDB pour la gestion centralisée des métadonnées et la supervision des flux permettant l'intégration rapide de nouveaux clients et le suivi automatisé des pipelines.
  - Développement, optimisation et exécution des traitements PySpark sur Azure Databricks pour l'intégration incrémentale, le calcul d'indicateurs et la mise à jour automatisée des jeux de données.
  - Calcul automatisé des KPI de performance, de sécurité et d'usage des batteries, utilisés pour le suivi client, la maintenance prédictive et l'amélioration continue des modèles analytiques.
  - Intégration des pipelines de features et des modèles ML dans Databricks en collaboration avec les Data Scientists, pour automatiser la prédiction du RUL et le recalcul des indicateurs de performance batteries
  - Conception d'un modèle hiérarchique multi-dimensionnel (client, site, rack, module, cellule) garantissant le stockage incrémental et la restitution analytique des indicateurs dans PostgreSQL.
  - Mise en place d'un système d'alertes métier en temps quasi réel avec InfluxDB et Grafana pour détecter les dérives thermiques et électriques et notifier automatiquement les clients.
  - Supervision technique industrialisée avec Grafana, Prometheus et Azure Monitor assurant la traçabilité, la disponibilité et la performance des workflows Airflow et Databricks en production.
  - Mise en œuvre d'une chaîne CI/CD complète sous Azure DevOps intégrant les tests unitaires PySpark, les validations métier et la documentation automatique des workflows.

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

Koacher

Data Engineer -2 ans en alternance + 3 mois freelance.

Lyon

De septembre 2022 à décembre 2024
- Conception et automatisation de pipelines d'ingestion via Airflow et Cloud Functions pour collecter les données applicatives et IoT, avec gestion des dépendances, reprise sur incident et contrôles de qualité.
  - Structuration du Lakehouse sur GCS et BigQuery en zones Bronze, Silver et Gold, garantissant la traçabilité et la cohérence entre les sources.
  - Développement de modèles dbt du staging jusqu'aux marts avec intégration des SCD Type 2, des tests de qualité personnalisés et de la documentation générée dynamiquement.
  - Mise en place d'environnements dbt dédiés au développement et à la production, intégrés à une chaîne CI/CD analytique automatisée grâce à GitHub Actions et Cloud Build.
  - Création de macros et de packages dbt internes afin de standardiser les règles de transformation et les contrôles de qualité à l'échelle de la plateforme.
  - Optimisation des performances et de la consommation BigQuery par partitionnement, clustering et suivi des coûts, pour améliorer la vitesse d'exécution et la maîtrise budgétaire.
  - Supervision complète de la fraîcheur des données, de la latence et du taux d'échec des workflows à l'aide de Cloud Monitoring et de tableaux de bord opérationnels.
  - Renforcement de la gouvernance et de la sécurité grâce à la gestion des accès IAM, aux journaux d'audit et à l'intégration du lineage dbt dans le catalogue de données interne.
  - Restitution des jeux de données dans Looker Studio et Tableau, permettant aux équipes produit et sportives de piloter la performance athlète et l'engagement utilisateur à partir d'indicateurs fiables et unifiés.

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

### 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)