



# Djalil Salah-bey

## Data & Analytics Engineer | Expert ETL/ELT (AirFlow, dbt, GCP, Databricks)

Ingénieur Data spécialisé dans la modélisation analytique, l'orchestration et la mise en production de plateformes de données cloud.  
Conçoit et automatise des pipelines fiables et scalables, en intégrant les bonnes pratiques de DataOps, CI/CD et observabilité, pour fournir des datasets traçables et exploitables par les équipes Data Science et métier.

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

- **Langages & frameworks**  
Python (pandas, PySpark), SQL analytique, Bash
- **Data Platform**  
Databricks (Azure / GCP), Delta Lake, BigQuery, ADLS Gen2, GCS
- **Orchestration & CI/CD**  
Airflow, Databricks Jobs, GitHub Actions, Azure DevOps
- **Transformation & modélisation**  
dbt (tests, docs, SCD2), Kimball / Snowflake
- **Streaming & ingestion**  
Kafka, Auto Loader, Fivetran, APIs REST, SFTP
- **Stockage & bases**  
PostgreSQL, MongoDB, InfluxDB, Parquet / JSON / Avro
- **Monitoring & gouvernance**  
Grafana, Prometheus, Marquez / OpenLineage, Unity Catalog, RBAC
- **Infra & DevOps**  
Docker, Terraform, Kubernetes

### Soft Skills

Autonome  
Curieux  
Proactif  
Pédagogue

### Expériences professionnelles

- **PowerUp Technology** Paris  
Data Engineer ( Freelance ) De mars 2025 à octobre 2025  

Conception et mise en production d'une **plateforme analytique OLAP ClickHouse** pour le suivi et la prédiction de la performance des batteries industrielles avec refonte du modèle de données hiérarchique et automatisation des traitements analytiques temps réel.

  - **Architecture & Modélisation OLAP** : conception d'un modèle multidimensionnel complet (faits, dimensions, hiérarchies spatiales et temporelles), normalisation Snowflake et définition des règles de binning et d'agrégation pour des analyses drill-down rapides et précises.
  - **Implémentation & Performance** : développement d'une base analytique optimisée sur ClickHouse (partitionnement, indexation avancée, vues matérialisées) garantissant **des performances OLAP < 100 ms** et une **stabilité en production**.
  - **Ingestion & Orchestration** : création de pipelines ETL orchestrés sous **Airflow**, intégrés à **Azure Databricks** pour les traitements distribués, et modélisation des transformations analytiques sous **dbt** (tests, documentation, métriques métiers unifiées).
  - **DataOps & Production** : automatisation CI/CD (GitHub Actions, Terraform) et supervision des traitements analytiques en production via Grafana et Marquez, assurant **des déploiements fiables et observables**.
  - **Valorisation & Exposition** : publication d'**APIs analytiques** et de **visualisations interactives** (Heatmaps, KPIs) via **FastAPI** et **Power BI**, offrant une restitution claire et partagée des indicateurs de performance.

**Résultats** : réduction du temps de traitement de 1 h à 5 min, requêtes < 100 ms, adoption complète en production et industrialisation d'une plateforme modulaire multi-clients.  
**Stack** : Python, SQL, ClickHouse, Airflow, dbt, Azure Databricks, Docker, Terraform, Grafana, Marquez, FastAPI, Power BI, PostgreSQL, InfluxDB.
- **Koacher** Lyon  
Data Engineer ( Alternance ) De décembre 2022 à décembre 2024  

Conception et industrialisation d'une **plateforme analytique cloud sur GCP (BigQuery)** destinée à la **centralisation, la transformation et la valorisation** des données issues des applications de suivi sportif.

  - **Architecture & ingestion** : définition des flux de données depuis l'app mobile, APIs et bases transactionnelles avec ingestion automatisée GCS → BigQuery orchestrée par Airflow.
  - **Modélisation** : conception d'un modèle en trois couches sous dbt Cloud, suivant les principes Kimball avec normalisation des dimensions et calcul des métriques business clés.
  - **Orchestration & automatisation** : planification et surveillance des pipelines sous Airflow, gestion des dépendances, des SLA, des retries et des notifications automatisées.
  - **CI/CD & DataOps** : intégration continue via GitHub Actions, exécution automatique des tests dbt, validation de schémas, packaging et déploiement contrôlé vers les environnements de production.
  - **Observabilité & gouvernance** : suivi de la fraîcheur des données, supervision des performances BigQuery, documentation automatisée et gestion des accès via IAM GCP.
  - **MLOps & scalabilité** : préparation de datasets pour l'entraînement de modèles prédictifs (segmentation, recommandation d'entraînements) et mise en place de pipelines de features reproductibles.
  - **Visualisation & valorisation** : exposition des données dans Looker Studio et Metabase, création de KPIs dynamiques et tableaux de bord interactifs pour les équipes métier.

**Résultat** : Industrialisation complète de la chaîne analytique du tracking applicatif à la restitution BI.  
**Stack** : Python, SQL, Airflow (Cloud Composer), dbt Cloud, BigQuery, GCS, GitHub Actions, Looker Studio, Metabase, GCP (IAM, Monitoring), PostgreSQL, Firebase, Docker.

### Diplômes et formations

- IA School — Diplôme RNCP Niveau 7 Expert Data & Ingénierie de l'Intelligence Artificielle (2023-2025)
- IA School — Bachelor 3 Expert Data & Ingénierie de l'Intelligence Artificielle (2022-2023)
- Université de Lille — Licence Banque, Finance & Analyse Quantitative (2019-2022)