



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
Patient
Pédagogue

Expériences professionnelles

- PowerUp Technology** Paris
Data Engineer (Freelance) De mars 2025 à octobre 2025
Conception et mise en production d'une base de données OLAP ClickHouse dans le cadre de la **refonte complète du modèle BIM** et de la **structuration des Data Marts métiers** pour le **suivi des performances des batteries industrielles (BESS)**.
 - Architecture** : conception d'une base analytique hiérarchisée sur ClickHouse, modélisée en Snowflake avec tables dimensionnelles et factuelles normalisées.
 - Modélisation** : intégration et transformation des données sources, mise en place des relations hiérarchiques multi-niveaux, et définition des règles de binning et d'agrégation dans un modèle OLAP optimisé pour les analyses multidimensionnelles et le drill-down rapide.
 - Structuration physique** : définition des clés hiérarchiques, stratégies de partitionnement, indexation, vues matérialisées et ingestion incrémentale optimisée pour ClickHouse.
 - Pipelines & orchestration** : conception de pipelines ETL sous Airflow, gestion des dépendances, alertes SLA, retries et intégration native avec les jobs Azure Databricks pour les traitements distribués.
 - CI/CD & qualité logicielle** : automatisation des tests unitaires, des stress tests et des déploiements via GitHub Actions et Terraform.
 - Observabilité & lineage** : supervision complète via Grafana, Marquez et Unity Catalog (traçabilité des flux, fraîcheur, documentation automatique).
 - Exposition & valorisation** : publication d'APIs analytiques et de visualisations (Heatmaps, KPIs BESS) via FastAPI avec rafraîchissement quasi temps réel des données.**Résultats** : adoption totale en production, réduction du temps de traitement de 1h à 5 min, requêtes Heatmap <100 ms, et base modulaire multi-clients permettant l'onboarding automatisé.
Stack : Python, SQL, ClickHouse, Airflow, dbt, Azure Databricks, Docker, Terraform, GitHub Actions, Grafana, Marquez, Delta Lake, Unity Catalog, FastAPI, PostgreSQL .
- Koacher** Lyon
Data Engineer (Alternance) De décembre 2022 à décembre 2024
Conception et industrialisation d'une **plateforme analytique cloud sur 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 les applications mobiles, APIs et bases transactionnelles (PostgreSQL, Firebase), ingestion automatisée vers **GCS → BigQuery**.
 - Modélisation** : conception d'un **modèle en trois couches (staging → intermediate → marts)** sous **dbt** suivant les principes Kimball avec normalisation des dimensions (utilisateur, session, activité, performance) et calcul des métriques business clés (rétention, engagement, churn).
 - Orchestration & automatisation** : mise en place de pipelines planifiés via **Airflow**, avec gestion des dépendances, SLA, notifications et reprise automatique.
 - CI/CD & DataOps** : intégration continue via **GitHub Actions**, tests automatiques (dbt + pytest), validation de schémas, packaging et déploiement contrôlé vers les environnements de prod.
 - Observabilité & gouvernance** : suivi de la fraîcheur des données, **tests de qualité dbt**, documentation automatisée et gestion des accès via **IAM GCP**.
 - MLOps & scalabilité** : préparation de datasets pour l'entraînement de modèles prédictifs (recommandation d'entraînements, segmentation utilisateurs) et mise en place de **pipelines de features** reproductibles.
 - Visualisation & valorisation** : exposition des données dans **Looker Studio**, création de KPIs temps réel et tableaux de bord interactifs pour les équipes métier et coaching.**Stack** : Python, SQL, GCP, BigQuery, Airflow, dbt, GitHub Actions, Looker Studio, PostgreSQL, Docker.

Diplômes et formations

- IA School — Master 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)