



Djalil Salah-bey

Data Engineer | Expert ETL/ELT (AirFlow, DBT)

Data Engineer déterminant sur des architectures Lakehouse et DataOps, de la capture IoT jusqu'à la restitution BI et MLOps.
Conçoit, industrialise et supervise des pipelines scalables sous Azure Databricks, dbt et Airflow, avec un haut niveau de fiabilité, traçabilité et automatisation CI/CD.

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

Passion

Sport
Programmation

Compétences Techniques

Langages : Python · SQL · PySpark · Bash
Orchestration & DataOps : Airflow · Azure Data Factory · Databricks Jobs · CI/CD
Transformation & Modélisation : dbt · Kimball · Snowflake · Lakehouse
Stockage & Cloud : Azure Data Lake Gen2 · BigQuery · ClickHouse
Monitoring & Observabilité : Great Expectations · OpenLineage · Grafana · Prometheus · Unity Catalog
Infra & DevOps : Docker · Terraform · Kubernetes · Azure Key Vault · IAM · Logging & Lineage
Visualisation & BI : Power BI · Looker Studio · Metabase

Soft Skills

- Curieux
- Proactif
- Rigoureux

Diplômes et formations

- IA School** — Diplôme RNCP Niveau 7 Expert Data & Ingénierie de l'Intelligence Artificielle (2022-2024)
- IA School** — Bachelor 3 Expert Data & Ingénierie de l'Intelligence Artificielle (2021-2022)
- Université de Lille** — Licence Banque, Finance (2018-2021)

Expériences professionnelles

●

PowerUp Technology

Freelance Data Engineer

Paris

De mars 2025 à octobre 2025

- Conception d'une **nouvelle architecture Lakehouse** sur **Azure Databricks** avec zones **Bronze → Silver → Gold** pour la gestion et l'historisation des données capteurs.
- Définition du **modèle analytique OLAP ClickHouse** : faits, dimensions et hiérarchies spatiales / temporelles adaptées aux heatmaps 2D/3D.
- Développement de **pipelines d'ingestion et de transformation PySpark / SQL** pour traiter les mesures IoT à grande échelle (plusieurs centaines de millions d'enregistrements).
- Orchestration des traitements sous **Airflow (AKS)** avec gestion des dépendances, SLA, alertes et reprise automatique en cas d'échec.
- Mise en place d'un **workflow CI/CD complet (GitHub Actions + Terraform)** : validation, tests dbt, déploiement automatisé multi-environnements.
- Intégration d'une **surveillance complète de la qualité et du lineage** :
 - Great Expectations** pour le contrôle des données,
 - OpenLineage / Purview** pour la traçabilité des flux,
 - Grafana / Prometheus** pour le suivi des SLA, latences et coûts.
- Sécurisation et gouvernance via **Unity Catalog** et **Azure Key Vault (IAM, gestion des secrets)**.
- Optimisation du traitement Spark et des vues ClickHouse : **requêtes < 100 ms, -30 % coûts compute, zéro rupture pipeline**.
- Collaboration étroite avec les équipes Data & R&D pour intégrer les analyses thermiques en production.

Résultats : Plateforme plus rapide, stable et supervisée de bout en bout (SLA > 99,9 %), Heatmaps 2D/3D actualisées quasi en temps réel et Adoption complète de la plateforme par les équipes Data et R&D.

Stack : Azure Databricks · Delta Lake · ClickHouse · Airflow · dbt · Terraform · GitHub Actions · Great Expectations · OpenLineage · Grafana · Prometheus · Purview · Power BI · MLflow

●

Koacher

Alternant Data Engineer

Lyon

De décembre 2022 à décembre 2024

- Conception d'une **architecture Lakehouse GCP** (GCS → BigQuery → BI) structurée en zones **brute, transformée et analytique**.
- Modélisation **Kimball** (staging → intermediate → marts) et intégration de **dbt Cloud** pour automatiser les transformations, tests et documentation.
- Développement de **pipelines Airflow / Cloud Composer** orchestrant les traitements batch et temps réel.
- Intégration des flux **Pub/Sub → Dataflow → BigQuery** pour ingestion d'événements applicatifs et marketing.
- Mise en place d'une **chaîne CI/CD complète (GitHub Actions + Cloud Build)** : lint, tests dbt, déploiement multi-environnements.
- Supervision Grafana / Cloud Monitoring** : métriques SLA, coûts et fraîcheur des données.
- Gouvernance et traçabilité via **Data Catalog / Dataplex** pour centraliser les métadonnées et le lineage.
- Création des datasets "Gold Layer" pour les KPIs (rétention, churn, LTV, activation).
- Publication de tableaux de bord interactifs **Looker Studio / Power BI** automatisés.
- Optimisation BigQuery : **requêtes 12x plus rapides, -25 % de coûts, zéro rupture pipeline** sur 6 mois.
- Collaboration étroite avec les équipes Produit et Marketing pour fiabiliser les indicateurs.

Résultats : reporting automatisé (3h → 15min), réduction significative des coûts et adoption généralisée de la plateforme par les équipes métier.

Stack : GCP (BigQuery, GCS, Dataflow, Composer, IAM) · dbt Cloud · Airflow · Terraform · GitHub Actions · Cloud Build · Grafana · Looker Studio · Data Catalog / Dataplex · MLflow