



Djalil Salah-bey

Data Engineer | Expert ETL/ELT (AirFlow, dbt, BigQuery, Databricks)

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

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 : Python, SQL, PySpark, Bash
- Intégration : Airflow, dbt Cloud, Talend
- DataOps : Azure, GCP, Terraform, Docker, GitHub Actions, Git
- Visualisation : Power BI, Looker Studio, Metabase
- Observabilité : Great Expectations, OpenLineage, Grafana, Prometheus, Data Catalog
- Stockage & DataLake : Databricks, Delta Lake, Azure Data Lake Storage (ADLS), Google Cloud Storage (GCS)
- Bases de données & Moteurs analytiques : BigQuery, PostgreSQL, ClickHouse, DuckDB, MongoDB, InfluxDB

Soft Skills

Autonome

Curieux

Proactif

Pédagogue

Expériences professionnelles

- PowerUp Technology

Data Engineer (Freelance)

Paris

De mars 2025 à octobre 2025

 - Conception et déploiement technique d'une architecture Lakehouse sur Databricks, structurée en zones Bronze → Silver → Gold, garantissant la clarté, la scalabilité et la traçabilité du modèle.
 - Développement de pipelines d'ingestion et de transformation PySpark / SQL, automatisant le nettoyage, l'enrichissement et l'agrégation des données dans des tables Delta.
 - Mise en œuvre du format Delta Lake (ACID) assurant la fiabilité transactionnelle et l'optimisation des performances (Z-Order, partitionnement dynamique, autoscaling des clusters).
 - Orchestration des workflows sous Airflow, avec dépendances, SLA, alerting et supervision des exécutions Databricks Jobs.
 - Création de modèles dbt Cloud connectés à Databricks SQL, avec documentation, tests automatiques et CI/CD GitHub Actions.
 - Implémentation d'une data observability avancée (Great Expectations, OpenLineage, Grafana) garantissant la qualité et le lineage de bout en bout.
 - Contribution active à la mise en place de la gouvernance et de la sécurité des données via Unity Catalog et Terraform, en coordination avec l'équipe infrastructure.
 - Publication des datasets "Gold" pour la BI, le Machine Learning et les analyses produits, en collaboration avec les équipes Data Science et métier.
 - Stack : Python, PySpark, SQL, Databricks, Delta Lake, Azure Data Lake Storage, Airflow, dbt Cloud, Great Expectations, OpenLineage, GitHub Actions, Grafana, Terraform, Power BI, Docker, Git.
- Koacher

Data Engineer (Alternance)

Lyon

De décembre 2022 à décembre 2024

 - Modélisation complète du Data Warehouse sous BigQuery selon les principes Kimball (staging → intermediate → marts), garantissant la cohérence et la standardisation des modèles métiers.
 - Développement et versionning des transformations SQL sous dbt Cloud, avec tests automatiques, documentation et revue de code intégrés dans la CI/CD.
 - Implémentation d'une gestion historique (SCD Type 2) pour suivre l'évolution temporelle des entités métiers clés et préserver la cohérence analytique.
 - Développement de pipelines Airflow et maintenance de flux Talend pour l'ingestion quotidienne de données issues de systèmes applicatifs et bases legacy (SQL Server, Oracle) vers BigQuery.
 - Automatisation et orchestration des workflows via Airflow & dbt Cloud Scheduler, avec gestion des dépendances, SLA, notifications et suivi d'exécution.
 - Documentation centralisée et gouvernance des métriques dans dbt Docs et Data Catalog, garantissant transparence, fiabilité et collaboration inter-équipes.
 - Intégration des datasets "Gold Layer" dans les outils BI pour le reporting automatisé, la mesure des KPIs et la prise de décision stratégique.
 - Contribution à l'industrialisation du cycle analytique via GitHub Actions et Terraform, avec tests unitaires et déploiement automatisé des modèles dbt en dev/preprod/prod.
 - Participation à l'optimisation des performances analytiques BigQuery (partitionnement, clustering, vues matérialisées, caching sélectif).
 - Stack : SQL, Python, BigQuery, SQL Server, Oracle, dbt Cloud, Talend, Airflow, Google Cloud Storage, Data Catalog, GitHub Actions, Grafana, Looker Studio, Power BI, Docker, Terraform, Git.

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)