



Djalil Salah-bey

Data Engineer | Analytics Engineer

Data Engineer avec 3 ans d'expérience sur des projets data de bout en bout : intégration de données (Talend, Python), plateformes cloud (Azure Databricks, Snowflake), orchestration (Airflow) et modélisation analytique (dbt, SQL). Habitué à travailler avec les équipes métier, produit et R&D pour transformer les besoins en indicateurs fiables et dashboards. En recherche d'un CDI pour continuer à construire et exploiter des plateformes data robustes.

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

Diplômes et formations

- IA School —Paris Master Ingénierie des données & de l'intelligence artificielle (RNCP Niv.7, 2022 – 2024)
- IA School —Paris Licence 3 Ingénierie des données & de l'intelligence artificielle (2021-2022)
- Université de Lille — Licence Banque, Finance (2018-2021)

Compétences Techniques

Cloud & Data Platforms

- Azure (Databricks, ADLS, AKS)
- GCP (GKE, Firebase, BigQuery)
- Snowflake, Databricks

Orchestration & Scheduling

- Apache Airflow (AKS, GKE)
- Databricks Workflows

Modélisation &

Transformation de données

- dbt Core
- SQL, PySpark, Pandas

Bases de données & Analytics

- PostgreSQL, MongoDB, DuckDB, ClickHouse
- Looker Studio, Power BI

Monitoring & Observabilité

- Grafana, Prometheus
- OpenLineage, Marquez

Développement &

Conteneurisation

- Python, Bash, Git
- Docker, Kubernetes, Linux

Expériences professionnelles

- **PowerUp Technology** Paris
De février 2025 à octobre 2025

Data Engineer - 9 mois Freelance

Environnement technique : Python, PySpark, Azure Databricks, ADLS Gen2, Git (Azure DevOps), Airflow sur AKS, Docker, PostgreSQL, MongoDB, ClickHouse

 - Conçu et développé des **pipelines batch de bout en bout** (Databricks / Airflow sur AKS) pour ingérer la télémétrie batterie (SFTP, APIs) vers ADLS Gen2, puis alimenter les tables analytiques utilisées par les dashboards clients.
 - Écrit et maintenu des **jobs PySpark** pour le nettoyage, la normalisation, les jointures avec les référentiels (clients, assets) et les **chargements incrémentaux** dans les différentes zones du Lakehouse.
 - Structuré le code en **modules Python réutilisables** (packages, fonctions utilitaires, configuration centralisée) et participé aux revues de code pour garantir lisibilité, testabilité et cohérence des pipelines.
 - Contribué à l'intégration d'**algorithmes et modèles Python** de l'équipe R&D dans les librairies internes, avec tests unitaires, tests d'intégration et déploiement via Azure DevOps.
 - Assuré le **run quotidien** des traitements : suivi des DAGs Airflow, analyse des logs, reprocess en cas d'incident et surveillance de la qualité des données (volumes, doublons, valeurs manquantes).
 - Collaboré avec les équipes **R&D et métiers** pour adapter les indicateurs calculés, les formats de sortie et les traitements aux contraintes opérationnelles des clients finaux.
- **Koacher** Lyon
D'octobre 2022 à octobre 2024

Data Engineer · Contrat d'alternance

Environnement technique : Python, Snowflake, dbt Core, Apache Airflow, SQL, APIs (Stripe, Firebase, HubSpot), Grafana, Git, Docker, Kubernetes

 - Développé des **pipelines d'ingestion Python orchestrés par Airflow** pour charger dans Snowflake les événements produit (Firebase), les paiements (Stripe) et les données CRM (HubSpot / systèmes internes).
 - Construit un **modèle de données Snowflake en couches** (staging, core, faits / dimensions) pour le suivi de l'usage, du funnel et des abonnements.
 - Implémenté les **transformations SQL dans dbt Core** (modèles, tests, documentation) pour produire des datamarts orientés abonnements, usage et revenus, avec modèles incrémentaux sur les tables volumineuses (events, paiements, abonnements).
 - Programmés des **DAGs quotidiens Airflow** pour rafraîchir les datamarts et les jeux de données utilisés par les dashboards métier et les analyses récurrentes.
 - Optimisé la **structure des tables et les requêtes Snowflake** afin de réduire les temps de réponse et la consommation (filtres, jointures, organisation des tables).
 - Défini des **conventions de nommage et de documentation** dans dbt (schéma, dossiers, descriptions) pour garder un modèle lisible, stable et facilement maintenable
- **Multi By EasyPlanet** Paris
De 2021 à 2022

Développeur ETL · Contrat d'alternance

Environnement technique : Python, Talend Data Integration, Talend ESB, PostgreSQL, SQL SERVER, ORACLE DB, SQL, Python, Linux, Git

 - Développé des **jobs Talend** pour ingérer dans PostgreSQL les données des applications métier, des fichiers plats et des APIs (SQL Server, Oracle, autres systèmes).
 - Mis en place des **flux ETL initiaux puis incrémentaux** pour alimenter un data mart opérationnel utilisé pour le reporting et le pilotage.
 - Configuré les **mappings Talend** : jointures, lookups, enrichissements, normalisation de valeurs, gestion des référentiels et tables de correspondance.
 - Construit des **vues et tables agrégées** dans PostgreSQL (SQL) pour exposer des indicateurs prêts à l'emploi aux équipes métier.
 - Ajouté des **contrôles de qualité** dans les jobs Talend (doublons, règles métier, dates incohérentes) avec journalisation des erreurs et gestion des rejets.
 - Écrit des **scripts Python** pour préparer les fichiers, automatiser certaines tâches récurrentes et contrôler la cohérence des jeux de données.
 - Surveillé l'**exécution des chaînes ETL** (planification, logs, reprise après incident) et documenté les flux pour faciliter la maintenance et les évolutions.