



Djalil Salah-bey

Data & Analytics Engineer | En recherche de CDI

Data Engineer expert en pipelines ETL/ELT automatisés, en modélisation analytique et en DataOps de production. Passionné par la programmation et l'automatisation, j'aime concevoir, déployer et fiabiliser des pipelines de données assurant la disponibilité, la traçabilité et la qualité des flux en environnement industriel.

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

Cloud & Data Platform

- Azure (Databricks, ADLS Gen2), GCP (BigQuery, GCS)

Orchestration & Scheduling

- Airflow (AKS, GKE), Databricks Jobs/Workflows

Modélisation & Transformation

- dbt (modèles métiers, tests, documentation)
- SQL, PySpark, Pandas

DataOps & Qualité

- GitHub Actions (CI/CD), Terraform (IaC)
- dbt tests, Pytest, Great Expectations

Monitoring & Observabilité

- Grafana, Prometheus, OpenLineage, Marquez

Databases & BI

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

Développement & Conteneurisation

- Python, SQL, Bash
- Docker, Kubernetes (minikube, AKS)

Diplômes et formations

- IA School —Paris Licence 3 & Mastère Expert en Data Engineering & Intelligence Artificielle (RNCP Niv.7, 2022 – 2025)

- IAE de Lille— Master Système d'information (2021-2022)

- Université de Lille — Licence Banque, Finance (2018-2021)

Expériences professionnelles

PowerUp Technology

Freelance Data Engineer -9 mois

Paris

De février 2025 à octobre 2025

- Conception, déploiement et fiabilisation de pipelines ETL/ELT** pour la supervision des batteries Battery Energy Storage Systems sous **Azure Databricks** en lien avec les équipes R&D, Produit et commerciales.
- Développement de pipelines PySpark et modules Python réutilisables**, versionnés sous Git, assurant l'ingestion, la transformation et la mise à disposition des données clients (SFTP → ADLS Gen2 → Databricks, structuration en zones Bronze / Silver / Gold) dans une logique Data Lakehouse.
- Industrialisation et mise en production des algorithmes et librairies analytiques** via Azure DevOps (Repos & Pipelines), intégrant tests unitaires, validations d'intégration et contrôles de qualité automatisés.
- Orchestration et planification des workflows sous Airflow (AKS)** garantissant la fiabilité, la traçabilité et la planification quotidienne des traitements distribués.
- Suivi opérationnel et supervision en production via Azure Monitor & Application Insights**, avec suivi des SLA, du taux de réussite et des volumes traités, garantissant la stabilité et la qualité des traitements.
- Analyses et visualisations exploratoires sur notebooks Databricks** en soutien aux équipes R&D, pour l'étude du comportement batterie (température, tension, cycles, SOH, vieillissement) et la définition d'indicateurs de performance, sécurité et vieillissement.
- Participation à l'amélioration continue du framework interne** : modularisation du code, documentation technique, bonnes pratiques Git et optimisation des traitements PySpark pour réduire coûts et temps d'exécution.

Stack : Python · PySpark · Azure Databricks · Airflow · Azure Kubernetes Services · ADLS Gen2 · Azure DevOps · Delta Lake · Git · Azure Monitor · Application Insights · Docker · PostgreSQL · MongoDB · ClickHouse · Azure Wiki

Koacher

Alternant Data Engineer - 2ans

Lyon

De octobre 2022 à octobre 2024

- Conception, évolution et maintenance des modèles de données dans dbt**, avec mise en place puis enrichissement d'une couche sémantique unifiée exposée aux utilisateurs métiers et aux équipes BI.
- Construction puis industrialisation progressive de pipelines ELT sous Airflow**, assurant la collecte, la transformation et le chargement automatisé de données multi-sources (APIs, SFTP, bases SQL, fichiers Cloud Storage) avec respect des SLA de traitement quotidiens.
- Rédactions des spécifications techniques** et de la modélisation fonctionnelle garantissant la qualité, la traçabilité et la cohérence des jeux de données livrés au fil des itérations.
- Généralisation des tests** (dbt tests, Pytest) et mise en place d'une chaîne CI/CD GitLab (lint, tests, déploiement automatisé) pour renforcer durablement la fiabilité, la maintenabilité et la revue de code des pipelines.
- Mise en œuvre puis amélioration continue des performances** (optimisation des requêtes SQL, maîtrise des coûts BigQuery, parallélisation des DAGs Airflow) avec réduction progressive des temps de rafraîchissement des principaux datasets.
- Implémentation et enrichissement de dashboards de suivi** (Airflow Metrics, Prometheus, Grafana) pour monitorer fraîcheur, volumétrie, erreurs et SLA, avec alerting proactif et baisse du nombre d'incidents en production sur la durée.
- Documentation et gouvernance des données** : définition et mise à jour des standards (naming, lineage, accès), rédaction des livrables techniques (schémas, dictionnaire de données, runbooks) et gestion continue de la couche sémantique dbt.
- Support technique et transfert de connaissances** auprès des équipes BI et Data Analysts sur l'usage des modèles dbt / SQL et l'exploitation des datasets BigQuery en self-service (accompagnement dans le temps sur les nouveaux besoins).

Stack : GCP (BigQuery, Cloud Composer, Cloud Storage) · Airflow · dbt · SQL · Python · GitLab CI/CD · Grafana · Prometheus · Talend · PostgreSQL · MongoDB

Projet personnel -Octobre 2025

- Pipeline DataOps autonome entièrement automatisé** de bout en bout sous Airflow (des appels tous les jours des APIs météo / pollution / trafic jusqu'à l'actualisation des dashboards Grafana), avec traitements PySpark et modélisation analytique sous dbt / DuckDB, déployé en environnement conteneurisé proche production (Docker / k8s).
- Stack :** Airflow · PySpark · dbt · DuckDB · Grafana · Marquez · GitHub Actions · Docker · Kubernetes
- Code :** github.com/IADJALILProject/iot-smartcity-data-platform