



Djalil Salah-bey

Data Engineer | En recherche de CDI

Data Engineer 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

Diplômes et formations

- IA School —Paris Master
Data Science (RNCP Niv.7,
2022 – 2024)
- IAE de Lille— Master
Système d'information
(2021-2022)
- Université de Lille —
Licence Banque, Finance
(2018-2021)

Compétences Techniques

Cloud & Data Platform

- Azure , GCP , Databricks

Orchestration & Scheduling

- Airflow (AKS, GKE),
Databricks Workflows

Modélisation & Transformation

- dbt, SQL, PySpark, Pandas

Monitoring & Observabilité

- Grafana, Prometheus,
OpenLineage, Marquez

Databases & BI

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

Développement & Conteneurisation

- Python, SQL, Bash
- Docker, Kubernetes

Expériences professionnelles

PowerUp Technology

Freelance Data Engineer – 9 mois

Paris

De février 2025 à octobre 2025

- Conçu, développé et industrialisé un **pipeline de données complet sur Azure** pour le traitement des mesures télémétriques issues de batteries connectées (BESS).
- Développé des **jobs PySpark et des scripts Python modulaires** (extraction, validation, enrichissement, fenêtres temporelles), avec configuration par environnement, logs structurés et versionnement Git.
- Implémenté l'**ingestion automatisée** dans **ADLS Gen2** avec stockage **Parquet partitionné** (client / site / jour, compression Snappy), jobs idempotents et gestion des doublons.
- Conçu la **couche analytique dbt (Silver → Gold)** : modèles SQL incrémentaux, tables de faits/dimensions, agrégations métier et tests de qualité (not null, unique, relations) documentés.
- Orchestré et supervisé les **workflows Airflow (AKS)** : DAGs paramétrables, gestion des dépendances, retries, suivi des SLA et alertes (mail / Teams) en cas d'échec ou de dérive.
- Industrialisation **CI/CD sous Azure DevOps** : pipelines de build et de déploiement, exécution automatique des tests dbt, linting Python et validations avant merge vers les environnements d'intégration et de production.
- Préparé des **datasets et features pour les Data Scientists** : jeux de données agrégés, fenêtres glissantes et labels de performance / dégradation des batteries, exposés pour les notebooks et pipelines ML.
- Documenté et structuré les **bonnes pratiques de data engineering** : schémas de flux, conventions de nommage, runbooks d'incident, guidelines de revues de code et de gestion de configuration.

Stack principale : Azure (Databricks, ADLS Gen2) · PySpark · dbt · Airflow (AKS) · Azure DevOps (CI/CD) · Python / SQL · Terraform · ClickHouse · n8n · MongoDB · influxDB · PostgreSQL · Power BI

Koacher

Alternant Data Engineer – 2 ans

Lyon

D'octobre 2022 à octobre 2024

- Maintenu et fait évoluer la **plateforme data GCP (BigQuery, Cloud Composer)** dédiée au suivi de la performance des athlètes et des capteurs connectés (données entraînements, metrics physiques, historiques).
- Développé et optimisé des **requêtes SQL BigQuery** et des **modèles dbt (staging → marts)** pour structurer un schéma en étoile et fiabiliser les KPIs métier (charges d'entraînement, progression, rétention).
- Conçu et automatisé des **pipelines d'ingestion et de transformation** (Python / SQL) orchestrés par **Airflow** : ingestion planifiée des sources (APIs, fichiers), contrôles de qualité (complétude, doublons, anomalies) et historisation.
- Analysé et résolu les **incidents** sur les workflows de données : diagnostic via logs Airflow / BigQuery, correctifs de code, gestion des cas limites (valeurs manquantes, retards, changements de schéma) et documentation des post-mortems dans Notion / Confluence.
- Mis en place et maintenu des **pipelines GitLab CI/CD** pour dbt et les jobs Python : tests automatiques, checks de schémas, déploiements standardisés vers les environnements d'intégration et de production.
- Collaboré avec les **data analysts et data scientists** pour **préparer des datasets d'analyse et d'entraînement** (fenêtres temporelles, agrégations, labellisation) alimentant des modèles de performance / fatigue athlétique et des dashboards Power BI.
- Rédigé et mis à jour la **documentation technique et fonctionnelle** : mapping des sources, définitions d'indicateurs, contrats de données avec les équipes métier et guidelines d'utilisation des jeux de données.
- Appliqué les **principes fondamentaux du data engineering** : approche ELT vers BigQuery, jobs idempotents, chargements incrémentaux, modélisation en étoile et gestion fine des droits et rôles GCP.

Stack principale : GCP (BigQuery, Cloud Composer) · PostgreSQL · MariaDB · Airflow · dbt · Python / SQL · GitLab CI/CD · Grafana · Ansible · Power BI

Projet Personnel 2025 open-source

- Développé un **pipeline Lakehouse automatisé** (Airflow, scripts PySpark/Python modulaires, dbt, DuckDB) collectant et transformant **quotidiennement** les données météo, pollution et trafic et stockage **Parquet partitionné**
- Modélisé un **schéma en étoile** et calculé les **indicateurs** avec dbt sur DuckDB pour back-end.
- Intégré dans des **DAGs Airflow réutilisables** l'ensemble des workflows d'ingestion, de transformation et de monitoring (Grafana / OpenLineage), packagés sous **Docker Compose** et exécutables sur **Kubernetes (minikube)**.
- Industrialisation **CI/CD via Makefile & GitHub Actions** (tests dbt, lint Python, build images) ; **résultat** : **pipeline totalement autonome**, orchestré chaque jour par Airflow avec **alertes e-mail en cas d'échec ou de dérive SLA**.

Stack : Airflow · PySpark / Python · dbt · DuckDB / Parquet · Docker / Kubernetes (minikube) · Makefile · GitHub Actions · Grafana / OpenLineage

Lien : <https://github.com/IADJALILProject/iot-smartcity-data-platform>