



Djalil Salah-bey

Data & Analytics Engineer (PySpark, Python, SQL)

Data Engineer passionné par la programmation et l'automatisation des pipelines de données.
J'aime coder chaque étape : ingestion, transformation, orchestration et supervision des flux.

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 & Plateformes :**
Azure (ADLS, ADF, Databricks, Purview), GCP (BigQuery, Dataflow, Cloud Composer), Docker, AKS/GKE
- Traitements :**
Spark/PySpark, SQL, dbt (tests/docs), Delta Lake, Parquet, ClickHouse, Kafka
- Orchestration & Qualité :**
Airflow, Databricks Workflows, data contracts, Great Expectations
- Observabilité & FinOps :**
logs/métriques/traces, Prometheus/Grafana, coûts & sizing
- Sécurité & RGPD :**
RBAC/IAM, secrets, masquage/tokenisation PII, rétention & purge
- CI/CD & IaC :** GitHub Actions, Azure DevOps, Terraform, tests unitaires & de régression
- BI/Consommation :** Power BI, Looker Studio, APIs/exports contractuels
- Langages :** Python (PySpark), SQL, Shell

Passions

Sport
Programmation
Intelligence Artificielle

Expériences professionnelles

- PowerUp Technology

Data Engineer

Paris

De février 2025 à septembre 2025

Freelance sur Azure Databricks, chargé de l'industrialisation des pipelines et de la fiabilisation des données BESS multi-clients.

 - Conception et orchestration** de pipelines batch sous **Airflow** collectant les fichiers de mesures BESS déposés sur **SFTP** clients vers **Azure Data Lake Gen2**, avec gestion des dépendances, logs et reprises automatiques.
 - Développement et orchestration** des traitements **PySpark** sur **Azure Databricks**, exécutés via **Databricks Jobs / Workflows** pour l'intégration incrémentale et le calcul automatisé des indicateurs.
 - Préprocessing automatisé des séries temporelles** avec étapes de filtrage, formatage, interpolation, complétion et validation de la consistance **des données capteurs**.
 - Calcul des KPI et alertes opérationnelles** (surchauffe, déséquilibre, sur-courant) et estimation du **RUL (Remaining Useful Life)** des modules batteries pour anticiper les défaillances.
 - Modélisation hiérarchique multi-clients** (client → site → rack → module → cellule) et stockage des indicateurs dans **PostgreSQL** (base de restitution) structurés par schéma client.
 - Développement d'une interface interne Python** pour la supervision et la gestion des métadonnées clients, appuyée sur **MongoDB** (métadonnées) et **InfluxDB** (séries temporelles).
 - Restitution continue** des données clients via des batches incrémentaux alimentant les schémas PostgreSQL consommés par les tableaux de bord front-end personnalisés.
 - Implémentation du Delta Lake** (time travel, schémas évolutifs, traitements incrémentaux) et **optimisation Databricks** (Z-Ordering, partitionnement, autoscaling).
 - Supervision technique et observabilité** des pipelines via **Grafana** (métriques Airflow/Databricks, taux de réussite, latence des jobs), consolidée avec **Azure Monitor** et **Log Analytics**.
 - Mise en place de tests unitaires PySpark** et validations métier garantissant la fiabilité des calculs avant publication et intégration **CI/CD Azure DevOps** pour le déploiement automatisé et la documentation continue.

Stack technique : Python · Spark(PySpark) · SQL · Azure Databricks · Airflow · Azure Data Lake Gen2 · Delta Lake · PySpark · PostgreSQL · MongoDB · InfluxDB · Grafana · Azure Monitor · Log Analytics · Azure DevOps · Docker · Git .
- Koacher

Data Engineer

Lyon

De décembre 2022 à décembre 2024

Participation à la conception et à l'industrialisation d'une plateforme analytique GCP centralisant les données sportives, applicatives et marketing.

 - Ingestion automatisée** des flux applicatifs (Firebase, Stripe, Hubspot, APIs partenaires) et capteurs sportifs via **Cloud Composer (Airflow)**, avec gestion des dépendances et contrôles de qualité.
 - Structuration du Lakehouse** en zones **Bronze / Silver / Gold** (GCS → BigQuery) afin d'assurer la traçabilité, la fiabilité et la cohérence des données inter-sources.
 - Modélisation analytique et transformation** sous **dbt + BigQuery** : création des dimensions hiérarchiques, tables de faits temporelles, gestion des SCD Type 2 et agrégations métier.
 - Développement de modèles dbt incrémentaux**, intégrant des tests automatiques (intégrité, cohérence métier) et la génération dynamique de documentation.
 - Orchestration des traitements dbt** dans **Airflow (Cloud Composer)**, avec notifications, logs et dépendances multi-jobs.
 - Implémentation du CI/CD** via **GitHub Actions** et **Cloud Build** pour automatiser les tests, le déploiement et la documentation des modèles dbt.
 - Optimisation des performances BigQuery** : partitionnement temporel, clustering, tuning des coûts et suivi de la consommation par dataset.
 - Restitution des indicateurs** dans **Looker Studio** (ou Power BI) pour le pilotage des performances sportives et l'analyse des usages applicatifs.
 - Suivi de la qualité et de la conformité** : catalogage des jeux de données, documentation centralisée et mise en place de conventions de nommage unifiées.

Stack technique : Python · SQL · GCP · BigQuery · DBT · Airflow (Cloud Composer) · GCS · Snowflake · Talend · GitHub Actions · Cloud Build · Looker Studio · Docker · Kubernetes · Git .

Diplômes et formations

- IA School** — Diplôme RNCP Niveau 7 Expert Data & Ingénierie de l'Intelligence Artificielle (2023-2025)
- IA School** — Bachelor Expert Data & Ingénierie de l'Intelligence Artificielle (2022-2023)
- Université de Lille** — Licence Banque, Finance (2019-2022)