



✉ salahbeydjalil@gmail.com

🔗 <https://djo-project-portfolio.vercel.app/>

📍 France

📞 +33 6 11 27 91 53

Langues

Français

Anglais

Arabe

Espagnol

Centres d'intérêt

Sport

Tennis

Basket Ball

Football

Voyages

Nouvelles cultures

Tour du monde

Tech

Conférences Data

Veille IA

Meetups

Compétences

- **Data Platforms :** Databricks, Delta Lake, dbt, Unity Catalog
- **Data Engineering :** Spark (PySpark, SparkSQL), ETL/ELT, batch/streaming, Kafka
- **Databases :** PostgreSQL, MySQL, MongoDB, InfluxDB, ClickHouse
- **Cloud & DevOps :** Azure, GCP, Docker, Kubernetes (AKS/GKE), Terraform
- **Orchestration :** Airflow, Databricks Workflows, CI/CD (GitHub Actions, Azure DevOps)
- **Observabilité :** Prometheus, Grafana
- **BI & Apps :** Power BI, Streamlit, APIs Flask
- **ML/MLOps :** MLflow, scikit-learn, TensorFlow, PyTorch

Djalil Salah-bey

Data Engineer | Disponible ASAP

Je suis Data Engineer avec 3 ans d'expérience, spécialisé dans les data platforms cloud, en particulier Databricks Lakehouse et les solutions cloud natives.

Au-delà de la technique, j'aime concevoir des pipelines end-to-end fiables et mesurables, qui transforment la donnée brute en indicateurs clairs, optimisent les coûts et fournissent aux équipes des produits data prêts à l'usage (dashboards, datasets ML, APIs).

Expériences professionnelles

Data Engineer

De mars 2023 à octobre 2025 **PowerUp Technology -EnergyTech** Paris

- **Conception et déploiement d'une Data Platform sur Azure (Databricks + Delta Lake)** dédiée à la gestion, la performance et la prédiction de la durée de vie de batteries IoT — approche transposable à l'analyse de signaux physiologiques.
- **Développement et maintenance de pipelines d'ingestion** (Azure Data Factory, Azure Blob, Kafka) pour la collecte de données capteurs à haute fréquence.
- **Traitement et préparation des séries temporelles** (SparkSQL, PySpark, dbt) selon l'architecture Médaillon (Bronze/Silver/Gold) : fenêtrage, filtrage, agrégations, feature engineering et labellisation pour les modèles ML.
- **Entraînement et suivi de modèles d'apprentissage automatique et profond** (scikit-learn, TensorFlow, PyTorch) intégrés dans des pipelines MLOps sous Databricks avec **MLflow** pour le tracking, la versioning et la répliquabilité.
- **Mise en place de la qualité et conformité** : tests unitaires (dbt tests, Great Expectations), contrôle d'accès et gouvernance (Unity Catalog, RBAC), respect des bonnes pratiques **RGPD et traçabilité** complète des données et modèles.
- **Optimisation des performances** : partitionnement, Z-Ordering, caching et gestion de la parallélisation pour réduire les temps de calcul et d'inférence.
- **Orchestration et industrialisation des workflows** avec Databricks Workflows & Airflow (gestion des dépendances, SLA, retries, backfills) et CI/CD (Azure DevOps, Git, Docker, Kubernetes AKS).
- **Monitoring et observabilité** (Grafana, Datadog) pour assurer la fiabilité des pipelines et des modèles en production.
- **Collaboration interdisciplinaire** avec les Product Owners et Data Analysts pour livrer des produits data : **dashboards analytiques Power BI/Streamlit**, data marts et jeux de données ML prêts à l'emploi.
- **Stack technique** : Azure (ADF, Databricks, Blob, AKS, DevOps), Delta Lake, Spark (PySpark, SparkSQL), dbt, MLflow, Unity Catalog, Airflow, Kafka, Docker, Kubernetes, Grafana, Datadog, Git, CI/CD, Tensorflow, Sckit-Learn, Keras.

Data Engineer

De décembre 2022 à novembre 2023 **Koacher -SportTech** Lyon

- **Développement et maintenance d'une plateforme SaaS de coaching sportif (web & mobile)** : architecture data, intégration applicative et automatisation des processus analytiques.
- **Conception et optimisation des bases de données relationnelles** (PostgreSQL, MySQL) et modélisation pour le reporting et la personnalisation des programmes utilisateurs.
- **Développement d'APIs REST sécurisées** (Python/Flask) pour la collecte et l'exploitation des données d'activité en temps réel.
- **Mise en place de pipelines ETL** (Python, dbt) pour nettoyer, transformer et agréger les données comportementales et d'engagement.
- **Automatisation des workflows** avec Airflow (gestion des dépendances, planification, reprise sur incident).
- **Tests unitaires et d'intégration** (pytest, dbt tests) garantissant la fiabilité et la reproductibilité des pipelines.
- **Conteneurisation et déploiement** sur Kubernetes (GKE) avec CI/CD (GitHub Actions) pour la scalabilité et la livraison continue.
- **Implémentation de la sécurité et conformité** (authentification OAuth2, gestion des accès et logs utilisateurs).
- **Suivi des performances et observabilité** via Prometheus & Grafana (dashboards internes, alertes automatisées).
- **Stack technique** : GCP, Kubernetes (GKE), Python, Flask, PostgreSQL, MySQL, dbt, Airflow, Docker, Git, CI/CD, Prometheus, Grafana.

Diplômes et Formations

BAC + 5 Ingénierie & Science des données

De 2023 à 2025 **IA School** Aix-en-Provence

BAC +3 Ingénierie & Science des données

De 2022 à 2023 **IA SCHOOL** Paris