



Djalil Salah-bey

Data & Analytics Engineer | Recherche CDI

Passionné par la donnée et son impact, j'aime construire des pipelines fiables et scalables qui transforment la donnée brute en valeur métier. Mon approche : rigueur technique, sens de la collaboration et souci constant de la qualité et de la traçabilité.

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

Passions

Sport
Programmation

Compétences Techniques

Langues
Python · PySpark · SQL · Bash · Scala

Cloud & Plateformes
Microsoft Azure , Databricks , Google Cloud Platform · Snowflake

Bases de données
Oracle DB, SQL Server, MongoDB, InfluxDB, PostgreSQL

Orchestration & CI/CD
dbt Core · Airflow (AKS) · Azure Data Factory · GitHub Actions · Azure DevOps · Docker · Kubernetes · Terraform

Visualisation & Monitoring
Power BI · Prometheus · Grafana

Machine Learning & MLOps
MLflow · scikit-learn · XGBoost · LightGBM

Soft Skills

- Proactif
- Autonome
- Curieux

Expériences professionnelles

●

PowerUp Technology

Freelance Data & AI Engineer

Paris

De février 2025 à octobre 2025

- Mise en place de pipelines d'ingestion batch quotidiens et incrémentaux pour centraliser les données IoT des batteries industrielles BESS.
- Récupération des flux via API REST (télémétrie applicative) et transferts SFTP (fichiers capteurs et historiques clients), puis stockage dans Azure Data Lake Storage.
- Structuration du stockage sur ADLS Gen2 en zones Bronze / Silver / Gold, conventions de nommage, partitionnement date / client / device, gestion des schémas et tables Delta Lake.
- Traitement sous Azure Databricks (PySpark) : formatage, filtrage, pré- et post-cleansing, reconstruction de séries manquantes et contrôle de cohérence.
- Calcul des indicateurs analytiques SoC, SoH et RUL ainsi que des alertes pour la surveillance du parc batterie, stockées dans InfluxDB pour le suivi temps réel.
- Enrichissement des jeux de données avec les métadonnées clients et techniques sous MongoDB, et stockage des configurations et résultats analytiques dans PostgreSQL.
- Modélisation analytique sous dbt-Databricks, avec tests de qualité, documentation automatique et déploiement CI/CD via Azure DevOps.
- Mise en place d'une orchestration unifiée combinant Databricks Workflows (traitements internes) et Airflow (AKS) pour l'automatisation complète des pipelines, avec relance intelligente des tâches en erreur, gestion fine des dépendances inter-systèmes et monitoring proactif des performances en production.
- Déploiement de tableaux de bord temps réel, alertes proactives et indicateurs de performance (SLA, taux de succès, latence) pour garantir la fiabilité et la stabilité de la production.

Stack : Azure Databricks · PySpark · Delta Lake · Databricks Workflows · Airflow (AKS) · dbt · Azure DevOps · ADLS Gen2 · MongoDB · InfluxDB · PostgreSQL · API REST · SFTP · Prometheus · Grafana · Terraform · Wiki

●

Koacher

Alternant Data Engineer

Lyon

De décembre 2022 à décembre 2024

- Conception et mise en production d'une plateforme Lakehouse GCP (Bronze / Silver / Gold) centralisant les flux applicatifs et IoT (capteurs de performance, GPS, biométrie).
- Ingestion automatisée des données via Cloud Composer (Airflow) et Dataflow (Beam) : collecte multi-sources (APIs, Firebase, SFTP) vers Google Cloud Storage.
- Transformation et modélisation analytique sous dbt + BigQuery, selon la logique Kimball (dimensions hiérarchiques, faits temporels, SCD Type 2).
- Mise en place de contrôles qualité (tests dbt, Great Expectations) et de règles métier (valeurs aberrantes, cohérence temporelle, filtrage par type de session).
- Documentation et traçabilité complète du lineage via OpenLineage, dbt docs et GCP Data Catalog pour la conformité et l'auditabilité.
- Supervision et observabilité des DAGs (logs Stackdriver, alertes Grafana, retries automatiques), garantissant robustesse et reprise sans perte.
- Automatisation du déploiement CI/CD via GitHub Actions (lint, tests, build, doc, deploy) pour fiabiliser la mise en production des transformations dbt.
- Exposition des jeux de données Gold vers Looker Studio / Power BI et vers des APIs internes pour les équipes Data Science et produit (analyse de performance, RUL, scoring athlète).
- Collaboration étroite avec les Data Scientists pour le pré-traitement et la mise à disposition des features d'entraînement prédictif sous BigQuery ML.

Stack : GCP · BigQuery · dbt · Cloud Composer · Airflow · Dataflow · Apache Beam · GCS · Docker · Stackdriver · Grafana · GitHub Actions · Looker Studio · Great Expectations · OpenLineage

Diplômes et formations

- Microsoft Azure DP – Certification Data Engineer Associate (en cours)
- IA School — Diplôme RNCP Niveau 7 Expert Data & Ingénierie de l'Intelligence Artificielle (2022-2025)
- IA School — Bachelor Expert Data & Ingénierie de l'Intelligence Artificielle (2021-2022)
- Université de Lille — Licence Banque, Finance (2018-2022)