



Djalil Salah-bey

Data & Analytics Engineer | En recherche de CDI

Data Engineer passionné par la programmation et l'automatisation des pipelines de données, j'aime concevoir et industrialiser chaque étape pour garantir la fiabilité, la performance et la traçabilité des flux en production.

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

Intelligence Artificielle

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

Expériences professionnelles

● PowerUp Technology

Freelance Data Engineer - 9 mois

Paris

De février 2025 à octobre 2025

- Conception d'une plateforme Data Lakehouse Azure centralisant les données IoT de batteries industrielles (ADLS Gen2, Delta Lake, Databricks).
- Mise en place de pipelines d'ingestion multi-clients avec Airflow sur AKS (SFTP, bases opérationnelles, APIs) avec gestion des dépendances, reprise et journalisation.
- Développement de traitements PySpark pour nettoyer, enrichir et agréger les séries temporelles (niveau client → site → rack → module → cellule).
- Calcul automatisé des KPI de performance, sécurité et vieillissement utilisés par les équipes data science et les équipes métier.
- Intégration de pipelines de features et modèles ML dans Databricks pour industrialiser la prédiction du RUL et le recalculation des indicateurs.
- Conception d'un modèle hiérarchique standardisé garantissant un stockage incrémental et une restitution analytique robuste dans PostgreSQL.
- Mise en place d'alertes quasi temps réel avec InfluxDB et Grafana pour détecter dérives thermiques/électriques et notifier automatiquement les clients.
- Supervision et observabilité des workflows (Airflow, Databricks) via Grafana, Prometheus et Azure Monitor, avec suivi de la latence, des erreurs et de la disponibilité.
- Industrialisation via une chaîne CI/CD Azure DevOps (tests PySpark, validations métier, documentation automatique) pour sécuriser les mises en production.

Impact : réduction d'environ 60 % des temps de traitement (de ~1 h à < 25 min) et disponibilité > 99 % sur les pipelines critiques.

Stack : Python · Spark / PySpark · SQL · Azure Databricks · Azure Kubernetes Service (AKS) · Airflow · Azure Data Lake Gen2 · Delta Lake · PostgreSQL · MongoDB · InfluxDB · Grafana · Prometheus · Azure DevOps · Docker · Git

● Koacher

Data Engineer -2 ans en alternance + 3 mois freelance.

Lyon
De septembre 2022 à décembre 2024

- Conception et automatisation de pipelines d'ingestion (Cloud Composer / Airflow, Cloud Functions) pour les données applicatives et IoT, avec contrôles de qualité et reprise sur incident.
- Structuration du Lakehouse GCP sur GCS et BigQuery en zones Bronze / Silver / Gold pour assurer traçabilité, historisation et cohérence entre les sources.
- Développement d'un projet dbt complet (staging → marts) incluant modèles incrémentaux, gestion SCD Type 2, tests de qualité et documentation générée automatiquement.
- Mise en place d'environnements dbt (dev / test / prod) intégrés dans une chaîne CI/CD analytique (GitHub Actions, Cloud Build) pour fiabiliser les déploiements.
- Création de macros et packages dbt internes pour standardiser règles de transformation, conventions de nommage et contrôles de qualité à l'échelle de la plateforme.
- Optimisation des performances et coûts BigQuery (partitionnement, clustering, refonte de requêtes lourdes) pour réduire la latence et maîtriser le budget.
- Supervision de la fraîcheur des données, de la latence et des échecs de jobs via Cloud Monitoring et tableaux de bord opérationnels.
- Renforcement de la gouvernance et de la sécurité (IAM, journaux d'audit, lineage dbt intégré au catalogue interne).
- Restitution des données dans Looker Studio et Tableau, pour le pilotage de l'engagement utilisateur et de la performance athlète à partir d'indicateurs fiables et unifiés.

Impact : fiabilisation d'une vingtaine de tableaux de bord métier (échecs de pipelines divisés par ~3, coûts BigQuery réduits d'environ 20-25 %).

Stack : Python · SQL · dbt · Airflow (Cloud Composer) · BigQuery · GCS · Cloud Build · GitHub Actions · Looker Studio · Tableau · Terraform · Docker · Kubernetes · Git

Diplômes et formations

- IA School — Diplôme RNCP Niveau 7 Expert Data & Ingénierie de l'Intelligence Artificielle (2022-2024)
- IAE de Lille— Master Système d'information & management de la donnée (2021-2022)
- Université de Lille — Licence Banque, Finance (2018-2021)