

# Mickaël Rakotoarinivo

## Data Engineer | Data Analytics Engineer

Azure Databricks | Snowflake | Python | PySpark | SQL | Data Warehouse | Lakehouse

Dailiens, VD | +41 79 600 42 86 | mickael.rakotoa@gmail.com | Permis B suisse | Mobilité Lausanne et Genève | Disponible le 1er octobre 2026

## PROFIL

Data Engineer confirmé avec plus de cinq ans d'expérience dans la conception, l'industrialisation et l'exploitation de plateformes data, de pipelines de production et de jeux de données analytiques. Expérience pratique sur Azure Databricks, Snowflake, tables Delta, PySpark, SQL, Python et modèles Medallion/Kimball. Chez InPost, travail sur Unity Catalog, Change Data Feed, Auto CDC, SCD, snapshots, Kafka, TOML, Spark Declarative Pipelines et GitLab CI/CD. Habitué au debugging, aux contrôles qualité, aux tests et à la mise en production de datasets, en collaboration avec la Finance, les Data Analysts, les Data Consultants, les Data Scientists et les équipes IT.

## COMPÉTENCES CLÉS

**Data engineering** Python, SQL, PySpark, Spark SQL, TypeScript, ETL/ELT, intégration de sources, S3, Kafka, traitements batch et streaming.

**Plateformes et modélisation** Azure Databricks, Snowflake, Unity Catalog, Delta Lake, tables Delta, data lakes, Data Warehouse, datamarts, architecture Bronze/Silver/Gold, architecture médaille, Kimball, modèles dimensionnels, modèles en étoile, tables de faits et dimensions, SCD1/SCD2/SCD3, snapshots journaliers, hiérarchies, WBS, grain, business keys, agrégats Gold.

**Qualité et industrialisation** Tests unitaires et d'intégration, TDD, debugging, contrôles de complétude, d'unicité et de rapprochement, couverture, monitoring, observabilité, lineage, performance, GitLab CI/CD, Agile, DevOps, revues de code, documentation et standards.

**Optimisation** Partitionnement, OPTIMIZE, Z-ORDER, broadcast joins, optimisation de la modélisation, des jointures, filtres et agrégations, orchestration de tâches Snowflake.

**BI et analytics** Tableau, Power BI et Power BI Service (conception de rapports, publication, espaces de travail, actualisations et droits utilisateurs), Alteryx, SQL Server Reporting Services.

**Outils et collaboration** Git, GitHub, GitLab, GitLab CI/CD, Docker, Jira, Confluence, Atlan, Playwright, Vitest, Convex, Railway, TOML, JetBrains IntelliJ et PyCharm, VS Code.

## EXPÉRIENCE PROFESSIONNELLE

### Data Engineer / Analytics Engineer - Modernisation data finance et BI

API SA | Mission chez tl - Transports publics lausannois | Lausanne, Suisse | Depuis juillet 2026

- Concevoir et maintenir des pipelines ELT en SQL et Python sur Snowflake, depuis des données financières brutes issues de SAP jusqu'aux couches Bronze, Silver et Gold.
- Transformer quotidiennement environ dix procédures stockées brutes en tables Silver selon un modèle en étoile pour la direction financière et le contrôle de gestion.
- Modéliser les centres de coûts, natures comptables, projets et WBS, ainsi que le réalisé et les engagements en faits, dimensions, hiérarchies et agrégats Gold pour Tableau; documenter le grain, les règles de transformation et les flux.
- Développer un framework Python inspiré de dbt pour gérer les dépendances, compiler et exécuter les traitements Snowflake, publier les données dans Tableau et visualiser le lineage.
- Industrialiser la qualité avec des tests unitaires et d'intégration, des contrôles de complétude, d'unicité et de rapprochement, et un seuil de couverture de 85 %.
- Fiabiliser les traitements et améliorer les étapes d'extraction, de transformation et de mise à disposition des données pour Tableau; concevoir des dashboards d'analyse financière.
- Revoir la codebase Data pour améliorer sa maintenabilité et sa capacité à évoluer.
- Collaborer avec la Finance et les Data Scientists via GitLab pour transformer besoins et scripts en datasets et pipelines documentés, versionnés, testés et réutilisables.

### Data Engineer - Architecture, industrialisation et exploitation

Mondial Relay by InPost Group | France | Septembre 2022 - juin 2026

- Être le Data Engineer côté Mondial Relay pour les marchés BE, NE, FR et LU, en collaboration avec les Data Engineers internationaux du groupe InPost sur des projets de données transfrontaliers.
- Concevoir et industrialiser des pipelines batch et streaming en PySpark et SQL sur Azure Databricks et Delta Lake, avec modélisation Kimball et architecture Bronze/Silver/Gold.
- Construire le flux de données brutes de S3 vers Kafka puis Bronze pour transporter les données en streaming et alimenter le lakehouse.
- Configurer Unity Catalog, Change Data Feed et Auto CDC pour créer les modèles, les tables Delta et les flux d'ingestion.
- Appliquer la méthode Kimball avec des dimensions SCD1, SCD2 et SCD3 et des fact snapshots journaliers; le `parcel_journey_snapshot` suit le parcours d'un colis au grain `parcel_number`.
- Modéliser une dimension transporteur au grain `carrier_id`, historisée en SCD3, et une dimension client chargeur au grain `customer_id`, historisée en SCD2.

- Développer et configurer des Spark Declarative Pipelines à partir de configurations TOML, de Bronze vers Silver puis Gold, pour ingérer des données aux formats Delta, Parquet, JSON et CSV.
- Optimiser des workflows Databricks par la modélisation, les jointures, filtres, agrégations, le partitionnement, OPTIMIZE, Z-ORDER et les broadcast joins; certains traitements sont passés d'environ 2 heures à 20 minutes, soit près de 70 % de temps de calcul en moins.
- Migrer des données Snowflake vers Databricks avec plus de 10 tables Bronze et 30 tables de faits et de dimensions Silver.
- Exploiter des workloads Snowflake, orchestrer des tâches et contribuer à une baisse de 20 % des coûts d'utilisation des warehouses XL.
- Mettre en place le monitoring de la qualité et l'observabilité des pipelines, avec un gain d'environ 2 heures pendant le run matinal.
- Livrer environ 20 rapports Power BI pour les opérations, la relation client et le secteur Damage & Loss, utilisés par une trentaine de personnes; construire les modèles depuis les tables Delta Silver et Gold, puis gérer publication, espaces de travail, actualisations et droits utilisateurs dans Power BI Service.
- Analyser les performances linehaul, hub, first mile et last mile, identifier les causes d'écart entre agences et hubs et contribuer à faire progresser plusieurs indicateurs de 80 % à 97 %.
- Créer des dashboards Tableau et migrer des dashboards vers Power BI.
- Développer et debugger les pipelines GitLab CI/CD, écrire en TDD des tests unitaires et d'intégration, et contribuer aux méthodologies DevOps et Agile.
- Assurer l'astreinte et le support utilisateur, onboarder des Data Analysts et Data Consultants, et animer des workshops internationaux sur les usages et bonnes pratiques Databricks en tant que référent certifié pour le marché BeNeFrLu.

### Data Analyst - Automatisation et qualité des données

Hoist Finance | Secteur bancaire | France | Mai 2021 - août 2022

- Automatiser des workflows d'extraction et de contrôle qualité sur Alteryx, avec des alertes pour détecter et signaler les anomalies, réduisant de 40 % le traitement manuel.
- Développer et maintenir des pipelines ETL planifiés pour alimenter le reporting opérationnel et stratégique.
- Concevoir et maintenir des tableaux de bord et rapports via Tableau et SQL Server Reporting Services.
- Réaliser des analyses statistiques et formuler des recommandations fondées sur les données pour les équipes métier.

## PROJETS SÉLECTIONNÉS

### PriceHunter Basket Optimizer | *Projet propriétaire | Depuis juin 2026 | price-hunter.app*

- Construire une application B2B d'optimisation de paniers avec ingestion des prix fournisseurs par web scraping, normalisation des produits et moteur de répartition des lignes selon les offres disponibles.
- Concevoir le workflow métier Search > Compare > Basket > Send et exposer les données via un backend Convex.
- Stack: TypeScript, Node.js, Convex, Playwright, Next.js, Vitest et Railway. Code propriétaire fermé; démonstration possible en entretien.

### Bitcoin Analysis Dashboard | *Projet Data Engineering end-to-end | 2025*

- Construire une architecture médaillon Bronze/Silver/Gold avec Databricks et dbt, des tables Delta et une ingestion quotidienne de données OHLCV en Python et PySpark.
- Configurer les modèles et tables dans Unity Catalog; développer un moteur de calcul d'indicateurs techniques RSI, MACD et EMA et livrer un dashboard web temps réel en React/Next.js.
- Stack: Python, PySpark, dbt, Databricks, Delta Lake et React/Next.js.

## CERTIFICATIONS

**Databricks** Certified Data Engineer | Certified Data Analyst

## FORMATION

**Master en Data Sciences** Université de Lille | 2020 - 2022

Modélisation de données, ETL, BI, analyse statistique, machine learning et projets professionnalisants.

**Bachelor en économie et gestion** 2017 - 2020

**Baccalauréat** Lycée "Collège de France", Madagascar | 2017

## LANGUES

**Français** Bilingue

**Anglais** Courant

**Malgache** Langue maternelle

## INFORMATIONS COMPLÉMENTAIRES

**Profils** LinkedIn | GitHub

**Centres d'intérêt** Musculation, arts martiaux (aikido), éducation financière, investissement et système monétaire.