

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 | Swiss B permit | Mobility: Lausanne and Geneva

Available 2 October 2026 |

<https://mickael-rakotoa.com>



PROFILE

Experienced Data Engineer with over five years designing, deploying and operating data platforms, production pipelines and analytical datasets. Hands-on experience with Azure Databricks, Snowflake, Delta tables, PySpark, SQL, Python and Medallion/Kimball models. At InPost, worked with Unity Catalog, Change Data Feed, Auto CDC, SCD, snapshots, Kafka, TOML, Spark Declarative Pipelines and GitLab CI/CD. Experienced in debugging, quality checks, testing and deploying datasets, working with Finance, Data Analysts, Data Consultants, Data Scientists and IT teams.

CORE SKILLS

Data engineering Python, SQL, PySpark, Spark SQL, TypeScript, ETL/ELT, source integration, S3, Kafka, batch and streaming processing.

Platforms and modelling Azure Databricks, Snowflake, Unity Catalog, Delta Lake, Delta tables, data lakes, data warehouses, data marts, Bronze/Silver/Gold architecture, medallion architecture, Kimball, dimensional models, star schemas, fact and dimension tables, SCD1/SCD2/SCD3, daily snapshots, hierarchies, WBS, grain, business keys, Gold aggregates.

Quality and production delivery Unit and integration tests, TDD, debugging, completeness, uniqueness and reconciliation checks, coverage, monitoring, observability, lineage, performance, GitLab CI/CD, Agile, DevOps, code reviews, documentation and standards.

Optimisation Partitioning, OPTIMIZE, Z-ORDER, broadcast joins, optimisation of data modelling, joins, filters and aggregations, Snowflake task orchestration.

BI and analytics Tableau, Power BI and Power BI Service: report design, publishing, workspaces, refreshes and user permissions. Alteryx, SQL Server Reporting Services.

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

PROFESSIONAL EXPERIENCE

Data Engineer / Analytics Engineer - Finance and BI data modernisation

API SA | Assignment at tl - Transports publics lausannois | Lausanne, Switzerland | July 2026 - Present

- Design and maintain SQL and Python ELT pipelines on Snowflake, from raw financial data sourced from SAP to Bronze, Silver and Gold layers.
- Transform around ten raw stored procedures daily into Silver tables using a star schema for finance leadership and management accounting.
- Model cost centres, account categories, projects and WBS, actuals and commitments as facts, dimensions, hierarchies and Gold aggregates for Tableau; document grain, transformation rules and data flows.
- Develop a dbt-inspired Python framework to manage dependencies, compile and execute Snowflake jobs, publish data in Tableau and visualise lineage.
- Implement unit and integration tests, completeness, uniqueness and reconciliation checks, with an 85% test coverage threshold.
- Improve processing reliability and the extraction, transformation and delivery of data for Tableau; build financial analysis dashboards.
- Review the data codebase to improve maintainability and support future development.
- Work with Finance and Data Scientists through GitLab to turn requirements and scripts into documented, versioned, tested and reusable datasets and pipelines.

Data Engineer - Architecture, production delivery and operations

Mondial Relay by InPost Group | France | September 2022 - June 2026

- Worked as the Mondial Relay Data Engineer for the BeNeFrLu markets, collaborating with InPost's international Data Engineers on cross-border data projects.
- Designed and deployed batch and streaming pipelines in PySpark and SQL on Azure Databricks and Delta Lake, using Kimball modelling and Bronze/Silver/Gold architecture.
- Built the raw data flow from S3 through Kafka to Bronze to stream data into the lakehouse.
- Configured Unity Catalog, Change Data Feed and Auto CDC to create models, Delta tables and ingestion flows.
- Applied Kimball modelling with SCD1, SCD2 and SCD3 dimensions and daily fact snapshots; parcel_journey_snapshot tracks each parcel's journey at parcel_number grain.
- Modelled a carrier dimension at carrier_id grain with SCD3 history, and a shipper customer dimension at customer_id grain with SCD2 history.
- Developed and configured Spark Declarative Pipelines using TOML configurations, from Bronze to Silver and Gold, ingesting Delta, Parquet, JSON and CSV data.
- Optimised Databricks workflows through modelling, joins, filters, aggregations, partitioning, OPTIMIZE, Z-ORDER and broadcast joins; cut selected runtimes from around 2 hours to 20 minutes.
- Migrated data from Snowflake to Databricks, including over 10 Bronze tables and 30 Silver fact and dimension tables.

- Operated Snowflake workloads, orchestrated tasks and contributed to a 20% reduction in XL warehouse usage costs.
- Implemented data quality monitoring and pipeline observability, saving around 2 hours during the morning run.
- Delivered around 20 Power BI reports for operations, customer relations and Damage & Loss, used by around 30 people; built models from Silver and Gold Delta tables, then managed publishing, workspaces, refreshes and user permissions in Power BI Service.
- Analysed linehaul, hub, first-mile and last-mile performance, identified causes of differences between branches and hubs, and helped improve several indicators from 80% to 97%.
- Created Tableau dashboards and migrated dashboards to Power BI.
- Developed and debugged GitLab CI/CD pipelines, wrote unit and integration tests using TDD, and contributed to DevOps and Agile practices.
- Provided on-call and user support, onboarded Data Analysts and Data Consultants, and ran international workshops on Databricks usage and best practices as the certified contact for the BeNeFrLu market.

Data Analyst - Automation and data quality

Hoist Finance | Banking | France | May 2021 - August 2022

- Automated extraction and quality-control workflows in Alteryx, with alerts to detect and flag anomalies, reducing manual processing by 40%.
- Developed and maintained scheduled ETL pipelines for operational and strategic reporting.
- Designed and maintained dashboards and reports in Tableau and SQL Server Reporting Services.
- Performed statistical analyses and provided business teams with data-based recommendations.

SELECTED PROJECTS

PriceHunter Basket Optimizer | *Proprietary project* | Since June 2026 | *price-hunter.app*

- Built a B2B basket optimisation app with supplier price ingestion through web scraping, product normalisation and a line-item allocation engine based on available offers.
- Designed the Search > Compare > Basket > Send workflow and exposed data through a Convex backend.
- Stack: TypeScript, Node.js, Convex, Playwright, Next.js, Vitest and Railway. Closed-source proprietary code; demonstration available at interview.

Bitcoin Analysis Dashboard | *End-to-end data engineering project* | 2025

- Built a Bronze/Silver/Gold medallion architecture with Databricks and dbt, Delta tables and daily OHLCV data ingestion in Python and PySpark.
- Configured models and tables in Unity Catalog; developed a calculation engine for RSI, MACD and EMA technical indicators and delivered a real-time React/Next.js web dashboard.
- Stack: Python, PySpark, dbt, Databricks, Delta Lake and React/Next.js.

CERTIFICATIONS

Databricks Certified Data Engineer | Certified Data Analyst

EDUCATION

Master's degree in Data Science University of Lille | 2020 - 2022

Data modelling, ETL, BI, statistical analysis, machine learning and applied professional projects.

Bachelor's degree in Economics and Management 2017 - 2020

Baccalauréat Lycée Collège de France, Madagascar | 2017

LANGUAGES

French Bilingual

English Fluent

Malagasy Native

ADDITIONAL INFORMATION

Profiles LinkedIn | GitHub

Interests Strength training, martial arts including aikido, financial education, investing and the monetary system.