

MICHAEL GARCIA

Data Engineer | Ph.D. Rigor for Robust Architecture & Data Quality (dbt)

@ michael.garcia.73@gmail.com
🌐 michaelg-create.github.io/portfolio

📍 Aix-les-Bains, France, willing to relocate near Geneva
LinkedIn: michaelgarcia838383
🔗 michaelg-create



PROFILE

PhD Physicist leveraging 15+ years of **analytical rigor** to build and optimize **production data systems**. Expertise in modern **Data Engineering** practices (Airflow, dbt, Cloud-Native AWS) and software development. Strong problem-solver proven in international R&D environments, dedicated to **data quality** and robust architectures. Highly adaptable with a strong commitment to rapidly acquiring new skills (streaming, PL/SQL).

PROFESSIONAL EXPERIENCE

Data Engineer & Software Developer

Autonomous Projects & Consultancy Focus

⌚ 2024 - Present 📍 Aix-les-Bains, France

Bank Branch Footfall Analytics – Production Pipeline

- End-to-end system: FastAPI → Implementation of **ETL processes** using PySpark transformations → Parquet/DuckDB → Streamlit dashboard
- **Airflow Orchestration:** Real-time DAG + backfill DAG for fault tolerance (managing complex data flow dependencies)
- **Data Modelling:** Designed and implemented data structures optimized for OLAP (DuckDB)

Financial Sectoral Analysis – Cloud-Native AWS Data Hub

- Multi-source ETL: 3 APIs (Alpha Vantage, Yahoo Finance, FRED) → S3 → Redshift
- Infrastructure: Terraform IaC, Data Orchestration with **Airflow (MWAA)**, CloudWatch monitoring
- Data transformations: **dbt Core** for ELT, implementing **Star Schema** principles for analysis
- **Data Quality & Lineage:** Used **dbt automated testing (90%+ coverage)** to ensure data integrity and generate **data lineage** documentation.

Physics & Chemistry Teacher

Ministry of Education

⌚ 2013 - Present 📍 Savoie, France

Software Development & Digital Transformation

- Developed SaaS application for classroom planning (**used by 20+ teachers**): Python/Streamlit → pure HTML/JavaScript for **GDPR compliance** (client-side data processing)

Physics Researcher (PhD)

LIPhy – CNRS

⌚ 2009 - 2013 📍 Grenoble, France

Scientific Computing, Large Dataset Analysis & Validation

KEY COMPETENCIES

- Analytical Rigor (PhD)
- Data Quality Focus
- International Collab.& Leadership
- Adaptability & Learning
- Clean Code Principles

TECHNICAL SKILLS

Data Engineering Stack

- **Orchestration:** Airflow (MWAA), DAG complexity management
- **Transformation:** dbt (Core, Automated Testing), ETL/ELT pipeline implementation
- **Big Data:** PySpark, HDFS (Familiarity)
- **Cloud/IaC:** AWS (S3, Redshift), Terraform, Docker, Kubernetes

Data Modelling & Databases

- **Databases:** PostgreSQL (**Advanced**), Redshift, DuckDB (OLAP)
- **SQL:** Advanced SQL scripting, Query Optimization, Transaction Management
- **Modelling:** Star Schema, Normalisation, Data Mart design

Programming & Software Engineering

- **Core Languages:** Python (**Advanced**, Data Processing Focus), SQL, JavaScript
- **Web/APIs:** FastAPI, Flask, REST APIs
- **Practices/CI/CD:** Git/GitHub Actions, Agile, Clean Code (SOLID)

LANGUAGES

French

English



Professional working proficiency (100% English in PhD research collaboration)

EDUCATION

Data Engineering Intensive Training

Data-Upskilling (400h program)

⌚ 2024 📍 Savoie, France

Focus on production-ready skills: Git workflows, testing, CI/CD, clean code, data tools (Spark, Airflow, dbt), cloud deployment, SQL.

- Built and automated data processing pipelines for optical microscopy experiments:
 - Reduced analysis time by **50%** via optimized **ImageJ/IDL routines**, accelerating the experimental feedback loop.
 - Analysed complex video datasets, extracting 50K+ trajectories, requiring robust **data extraction and cleaning**.
 - Applied **scientific rigor** to implement simulation models for **cross-validation** against experimental results (directly comparable to software validation/testing).
 - **International collaboration:** 6+ researchers across 5 countries (100% English).

PROFESSIONAL DEVELOPMENT

Technical Deep Dive

Focus on Required Stack

2025

Savoie, France

- Advanced Data Modelling (Star Schema implementation, Lakehouse principles) and Relational Database deep-dive (Query Optimization, stored procedures concept).
 - Initiating PoC on **data streaming** and **messaging systems** (Kafka/RabbitMQ) to meet modern data platform requirements.

Ph.D. in Physics for Life Sciences

Université Grenoble Alpes

2009 - 2013

Grenoble, France

Optical microscopy, large-scale data analysis, scientific computing

M.Sc. Statistical Physics

École Normale Supérieure de Lyon

2005 – 2007

Lyon, France

B.Sc. & M.Sc. (1st year) in Physics

Université Savoie Mont Blanc

2001 – 2005

 Savoie, France

AVAILABILITY

Available for relocation to Geneva in one month.

REFERENCES

Available upon request.