

MICHAEL OZEH OKANG

Data Engineer | Analytics Engineer | Cloud Data Platforms | Streaming & ELT

Vienna, Austria | Okang.michael@gmail.com | +43 664 994 75957 | github.com/Matoxki | linkedin.com/in/michael-okang-ozeh

PROFESSIONAL SUMMARY

Project-based Data and Analytics Engineer with hands-on experience designing modern batch and streaming platforms that transform fragmented operational data into governed, analytics-ready products. Builds end-to-end ELT and event-driven architectures with Python, SQL, Apache Spark, Databricks, Kafka, Delta Lake, dbt, Airflow, Snowflake and BigQuery. Focused on reliable orchestration, data quality, historical tracking, governance and maintainable data products that support trusted business decisions.

TECHNICAL SKILLS

- **Programming & Processing:** Python, SQL, PySpark, Spark SQL, Apache Spark, Structured Streaming
- **Data Engineering & Lakehouse:** Databricks, Delta Lake, dbt, Medallion Architecture, ETL/ELT, incremental processing
- **Integration & Streaming:** Kafka, kafka-python, Confluent Cloud, Snowpipe, REST APIs, CDC, Debezium
- **Cloud, Platforms & Storage:** Snowflake, BigQuery, SQL Server, PostgreSQL
- **Orchestration, Governance & Delivery:** Airflow, Databricks Workflows, Unity Catalog, lineage, Docker, Git, GitHub Actions
- **Modelling, Analytics & Languages:** Kimball, star schema, SCD, Power BI, DAX; English (native), German (A2)

SELECTED DATA ENGINEERING PLATFORMS

PROJECT ALPINE

Real-Time Multi-Exchange Market Streaming & Lakehouse Platform

Python | Confluent Cloud Kafka | PySpark Structured Streaming | Databricks | Delta Lake | dbt
github.com/Matoxki/project-alpine

- Completed an end-to-end streaming lakehouse that ingests transaction-level Kraken and Coinbase trades through Python producers and Confluent Cloud Kafka, enforcing a consistent trade grain across sources.
- Implemented stateful deduplication, checkpointed Bronze/Silver Delta processing and a dbt dimensional layer with fact_market_trades, dim_asset, dim_exchange and dim_time.
- Validated ingestion and transformations locally with Docker before deploying to Databricks on AWS, orchestrating with Workflows and applying Unity Catalog basics to control trial compute costs.

PROJECT DANUBE

Enterprise CDC, History Tracking & Governed Data Platform

PostgreSQL | Debezium | Kafka | Databricks Auto CDC | Delta Lake | dbt | Unity Catalog
github.com/Matoxki | [repository to be published](#)

- Designing a governed CDC platform from PostgreSQL through Debezium and Kafka into immutable Bronze history, decoupling operational workloads from analytics.
- Using Databricks Auto CDC for deterministic insert, update and delete handling across Silver current-state and SCD history tables, followed by dbt Gold operational marts.
- Building reconciliation tests, idempotent backfills, lineage, governed access and threshold/webhook monitoring for reliable operations.

Real-Time E-Commerce Streaming & Cart-Abandonment Platform | Kafka | Snowflake | dbt | Airflow

github.com/Matoxki/Real-Time-ECCommerce-Streaming-Pipeline

- Streamed customer events into Snowflake and modelled near-real-time abandoned-cart leads with dbt; orchestrated tested micro-batches in containerized Airflow.

Vienna Transit Intelligence Platform | Python | BigQuery | dbt | Airflow

github.com/Matoxki/vienna-transit-pipeline

- Built an incremental ELT pipeline for Vienna transit and weather APIs with dependency-controlled Airflow DAGs, dbt models and quality/freshness tests.

DATA ANALYTICS & BI PROJECTS

Chocolate Sales Intelligence | Python | Power BI | DAX | Star Schema

github.com/Matoxki/Chocolate-Sales-Analytics-End-to-End-BI-Pipeline

- Cleaned and modelled 200,000 sales records into an analytical star schema; restored about 1,900 valid rows after diagnosing a transformation tolerance defect.

Cultural Heritage & Sustainability Analytics | SQL Server | Data Modelling | Power BI

github.com/Matoxki/Cultural-Heritage-Tourism-Data-Modeling-Sustainability-Analytics

- Standardized and modelled 15,000 tourism and sustainability records into a governed relational structure for cross-domain analysis and reporting.

PROFESSIONAL EXPERIENCE

Service Station Associate | Shell, Hof bei Salzburg, Austria

Feb 2025 - Feb 2026

- Managed point-of-sale transactions, cash reconciliation and operational records accurately under time pressure.
- Monitored inventory levels and reported discrepancies, supporting dependable daily supply and control logs.

EDUCATION

MSc, Sustainability Assessment and Resource Management (part-time)

Expected start Sep 2026

Hochschule Campus Wien - University of Applied Sciences

- Focus: sustainability reporting, life-cycle assessment and ESG/CSRD data governance.

MSc, Applied Geoinformatics (currently enrolled)

Mar 2025 - Present

University of Salzburg

- Coursework in geographic information science, spatial analysis, ArcGIS and PostgreSQL.

BSc, Urban and Regional Planning

Sep 2006 - Jan 2013

Olabisi Onabanjo University

- Grade: Second Class Upper

CERTIFICATIONS

- Mastering Advanced SQL Queries - Coursera, issued Jul 2026
- SQL Foundations - Microsoft (Coursera), issued Jul 2026
- Data Manipulation and Transactions in SQL Server - Microsoft (Coursera), issued Jul 2026
- Power BI Essentials and Advanced Techniques - Packt (Coursera), issued Jul 2026