Michael Muolokwu

Data Engineer | Data Scientist

Vancouver Canada muolokwunonso@gmail.com +17789576183 Michael Muolokwu Github

Summary

Google Cloud certified Data Engineer with over 4 years of experience in building and refining data-driven systems that bring complex ideas to life. Every day, I work on designing, deploying and optimizing pipelines, models, and architectures that extract insights and drive decisions internally and on client-facing projects. I thrive on breaking down intricate data challenges, solving tough debugging issues, and transforming raw information into actionable intelligence. I bring to work a relentless drive to create impact and a good team spirit to analyze, create, and innovate that fuels me, and I'm always ready to push boundaries, and shape the future of data-driven technology.

Skills and Technologies

- **Programming Languages:** SQL, Python, R, Golang
- **Cloud Services:** AWS, GCP, dbt
- **Orchestration Tools:** Apache Airflow, Prefect, PostgreSQL, Github Actions, GNU Make, CI/CD
- **Databases/Datawarehouses**: MySQL, PostgresSQL, Redshift, Bigquery, Databricks, Mongodb
- **Big Data**: Spark, Hive, Hadoop, Kafka

Education

University of British Columbia *Sept 2024 – June 2025*
MS in Data Science

- **Coursework:** Machine Learning, Data Structures and Algorithm, Deep Learning, Cloud Infrastructure

University of Nigeria *Oct 2015 – June 2020*
BEng in Electronics Engineering

- GPA: 4.70/5.0, Best Graduating Student in the Department
- **Coursework:** Digital Electronics, Communication Systems, Control Engineering, Software Engineering

Experience

Data Engineer *London, UK*
Luminr *Nov 2023 – Present*

- Developed and implemented a real-time processing pipeline using GCP's Pub/Sub and Dataflow, enabling the unboarding of new clients in real-time.
- Engineered and optimized over 5 ETL data pipelines on Airflow utilizing Python and SQL, leading to a 40% reduction in data processing time while ensuring seamless integration with GCP services for real-time analytics.
- Conducted comprehensive research on security frameworks, resulting in the implementation of best practices that decreased unauthorized access incidents by 70%, safeguarding sensitive information across a database exceeding 500GB.
- Led a team training session on emerging ELT frameworks, including dbt, enhancing the team's ability to evaluate and select the most effective tools for each project.
- Optimized SQL queries, making it more cost efficient by reducing the amount of data scanned in the data-warehouse and also reducing the query runtime by 30%
- Developed comprehensive documentation and automated testing frameworks for over 10 data workflows on Airflow, which hastened unboarding by 60% and minimized downtime across all systems.
- Optimized data quality protocols across a centralized data warehouse by developing automated validation scripts, leading to a 40% reduction in data discrepancies and enhancing reporting accuracy for over 150 business units.

Data Scientist *Remote*
ehealth4everyone *Jan 2022 – Nov 2023*

- Collaborated with cross-functional teams to define clear objectives and requirements, translating business needs into actionable machine-learning solutions.
- Collaborated with DevOps and infrastructure teams to ensure seamless integration of machine learning solutions with existing production systems increasing project completion time by 20%.
- Conducted A/B testing on survey data using bootstrapping, to provide statistical evidence for Stakeholder meetings.
- Conducted comprehensive code reviews for a team of 5 data scientists, enhancing adherence to best practices and reducing code-related errors by 40%, streamlining project delivery timelines.

Awards

UK IT Industry Awards finalists 2024 Nov 2024
Personal Excellence, Engineer of the Year
Engineer of the Year

Mastercard Scholarship Mar 2024
Mastercard Scholarship

Projects

Currency Converter *github.com/Converter*

- Developed a mid-market currency converter API using FastAPI. It had three different routes; Convert, Currency, and History
- All endpoints were protected with API key auth used in logging in.
- MongoDB was used in caching data history from the user and storing it in a cluster

Audio Tagging System *github.com/audio*

- Predicted audio classes using different models implemented using Keras. Models included a One-dimensional convolution, Two-dimensional convolution, and RNNs
- Tools Used: Python, Keras

Certifications

AWS Certified Cloud Practitioner July 2022
dbt Jan 2023
Professional Data Engineer Dec 2024
AWS ML Engineer Nanodegree Nov 2024
Machine Learning, Stanford Oct 2019