(+49) 157 39814757

Berlin, Germany

emyraeleson@gmail.com

GitHub: Emyrael

LinkedIn: Emmanuel-Onwubuya

Emmanuel Onwubuya

Data Engineer

SUMMARY

High-impact Data Engineer with 4+ years of experience building scalable data pipelines, designing resilient data models, and leading cloud-based data platform migrations. Adept at translating complex business needs into robust, production-grade data solutions that power analytics, machine learning, and real-time decision-making. Skilled in Python, SQL, and Spark, with hands-on experience using tools and platforms such as dbt, Databricks, Airflow and Terraform. Deep understanding of data modeling, orchestration, CI/CD, and cloud infrastructure (AWS, Azure, GCP). Currently exploring applications of AI and large language models (LLMs) to enhance data workflows, including natural language interfaces for querying structured and unstructured databases. Known for improving pipeline performance, ensuring data integrity, and bridging technical depth with business context to deliver measurable impact.

SKILLS

Programming & Scripting

Python, SQL, PySpark, Bash, JavaScript

Tools & Frameworks

Databricks, DBT, Supabase, Airflow, Terraform, Docker, Git, CI/CD, REST APIs

Data Platforms & Cloud

Azure, AWS, GCP, Snowflake, Databricks, Tableau, Looker, Power BI

Data Engineering & AI

ETL/ELT, Data Warehousing, Data Modeling (Bronze–Silver–Gold), Data Quality, Workflow Orchestration, Prompt Engineering, Model Context Protocol

Soft Skills

Analytical thinking, problem-solving, stakeholder management, agile collaboration, ethical judgment, teamwork

Languages

English (Native), German (Conversational), Russian (Conversational)

Interests

Scalable Data Systems, Cloud Infrastructure, Big Data Methodologies, Applied Machine Learning, Generative AI

EXPERIENCE

Data Engineer | Wefra Life

Frankfurt, Germany · 2024 – Present

- Migrated more than 200 ETL models from DBT to Databricks, improving pipeline performance and aligning architecture with business and analytics goals.
- Designed and managed Bronze–Silver–Gold data layers for ingestion, transformation and presentation, ensuring scalable, reliable and well-documented data models.
- Developed prompt-engineering techniques to optimize the performance and output of LLM-driven workflows within the data platform.
- Connected databases via Model Context Protocol (MCP) to enable natural language queries and LLM-powered database manipulation, accelerating analytics for business users.
- Delivered analysis-ready datasets to Tableau and media dashboards, supporting downstream analytics and data-driven decision-making.
- Automated Databricks workflow deployment using Terraform with comprehensive monitoring and alerting to handle schema changes and ensure data integrity.

Data Engineer | Accenture

Hamburg, Germany · 2022 – 2024

- Developed a cloud-native data warehouse by transforming large, heterogeneous datasets into scalable PySpark and SQL pipelines on GCP, providing unified data access.
- Collaborated with stakeholders across business, analytics and executive teams to align data products with strategic KPIs, ensuring adoption and maximizing business impact.
- Automated complex ETL workflows and scheduled data-building processes, improving data freshness and reducing manual processing efforts by over 30%.
- Optimized database performance through partitioning, indexing and refactoring query logic, reducing data load times and enhancing end-user responsiveness.

Junior Data Engineer | Domicil Real Estate Group

Munich, Germany · 2021 – 2022

- Automated recurring ETL workflows using Python and shell scripts, improving data processing efficiency and enabling faster delivery of business insights.
- Collaborated with the finance team to implement real-estate acquisition and sales models, enhancing flexibility and contributing to an increase in successful acquisitions.
- Leveraged tools such as Google BigQuery and DBT to streamline data extraction, transformation and modeling processes, improving analytics accuracy and reducing time-to-insight.

PROJECTS

Ufindar – AI-Powered University Search Tool

- Created Ufindar, an AI-enabled university search engine that allows users to discover universities using natural-language queries.
- Integrated LangChain with Supabase and embedding-based semantic search to map user questions to relevant universities, programs and metadata.
- Developed a backend with Python, Supabase and embeddings for efficient retrieval and deployed the prototype dev.ufindar.com.

EDUCATION

- Master in Data Analytics and Machine Learning, Universität Hildesheim, Germany
- Bachelor of Engineering in Information Systems and Technologies, Voronezh State University, Russia