

(+49) 157 39814757
Germany
emyraeleson@gmail.com

Emmanuel Onwubuya

Data Engineer

GitHub: Emyrael
LinkedIn: Emmanuel-Onwubuya
Portfolio: Projects

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, JavaScript, Bash, LLM Prompting
Tools & Frameworks	Airflow, Git, Terraform, MCP,Docker, CI/CD, REST APIs, Google Analytics
Data Platforms & Cloud	Databricks, DBT, Snowflake, AWS, GCP, Azure, Tableau, Supabase, Looker, Power BI
Data Management	ETL/ELT, Data Warehousing, Data Modeling, Data Quality, Database Management, A/B Testing.
Strengths	Analytical Thinking, Problem-Solving, Stakeholder Management, Agile Collaboration, Time Management, 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	2024 — Present Frankfurt, Germany
<ul style="list-style-type: none">• Migrated more than 200 ETL models from DBT to Databricks, significantly improving pipeline performance and aligning data architecture with evolving business and analytics requirements.• Designed and managed Bronze–Silver–Gold data layers for ingestion, transformation, and presentation—ensuring scalability, reliability, and comprehensive documentation of data models.• Developed advanced prompt-engineering techniques to optimize the performance and outputs of LLM-driven workflows embedded within the data platform.• Connected enterprise databases via Model Context Protocol (MCP) to enable natural language queries and LLM-powered database manipulation, accelerating analytics and empowering business users with self-service insights.• Delivered production-grade, analysis-ready datasets to Tableau and media dashboards, supporting downstream analytics, marketing insights, and data-driven decision-making.• Automated Databricks workflow deployment with Terraform, establishing robust monitoring and alerting mechanisms to handle schema drift, prevent failures, and safeguard data integrity across pipelines.	
Data Engineer <i>Accenture</i>	2022 —2024 <i>Hamburg, Germany</i>
<ul style="list-style-type: none">• Led the development of a robust data warehouse solution by transforming large-scale, heterogeneous datasets into scalable PySpark and SQL pipelines, enabling unified data access across departments.• Partnered with stakeholders across business, analytics, and executive teams to align data products with strategic KPIs, ensuring adoption and maximizing downstream business impact.• Automated complex ETL workflows and scheduled data building processes, improving data freshness, model readiness, and reducing manual processing efforts by over 30%.• Optimized database performance and query execution through partitioning, indexing, and refactoring logic, reducing data load times by up to 40% and enhancing end-user responsiveness.	
Junior Data Engineer <i>Domicil Real Estate Group</i>	2021 — 2022 <i>Munich, Germany</i>
<ul style="list-style-type: none">• Automated recurring ETL workflows using Python and shell scripts, improving data processing efficiency by 20% and enabling faster delivery of business insights to internal teams.• Collaborated with the finance team to implement real estate acquisition and sales models, enhancing flexibility and contributing to a 5% 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.	

EDUCATION

Masters in Data Analytics and Machine Learning, *Universität Hildesheim - Germany.*

Bachelor of Engineering: Information Systems And Technologies, *Voronezh State University Russia*

PROJECTS

Ufindar – AI-Powered University Search Tool

- Created **Ufindar**, an AI-powered university search engine enabling students to discover universities through natural-language queries rather than rigid filters.
- Integrated LangChain with Supabase and embedding-based semantic search to map user questions directly to relevant universities, programs, and metadata—delivering contextual, accurate results.
- Engineered the backend using Python, Supabase, and vector embeddings for efficient retrieval, and deployed the working prototype at dev.ufindar.com.