

E

EMEKA NWEKE EDEH

Data Engineer

CONTACT

Berlin, Germany
nweke.edeh@gmail.com
+49 176 88426042
[linkedin.com/in/edeh](https://www.linkedin.com/in/edeh)

CORE STACK

CLOUD
AWS, GCP, Azure, Terraform

DATA ENGINEERING
dbt, PySpark, Kafka, Airflow, Databricks

WAREHOUSES
Snowflake, BigQuery, Redshift

STREAMING & CDC
Kafka, Kinesis, DMS CDC

DEVOPS
GitHub Actions, Docker, CI/CD, OIDC

AI & APPLICATIONS
Claude API, FastAPI, Streamlit

LANGUAGES
Python, SQL, R

BI & ANALYTICS
Power BI, Tableau, Looker, QuickSight

LANGUAGES

English Native
German B1, developing

EDUCATION

B.Sc. Real Estate
Enugu State University of Science and
Technology

AUTHORISATION

EU work permit, no sponsorship required

BUILDING RELIABLE DATA PLATFORMS

Data engineer specialising in production lakehouse architectures, streaming pipelines, and infrastructure-as-code. Recently extended into LLM application development, shipping a production natural language analytics agent on top of the platform I built.

4+

YEARS EXPERIENCE

1M+

RECORDS DAILY

AWS • GCP

CLOUD PLATFORMS

PROFESSIONAL EXPERIENCE

Independent Data and Platform Engineer

2025 to Present

Consulting Practice Berlin, Germany

- ▶ Architected a production AWS medallion lakehouse: PostgreSQL via DMS CDC into Bronze, Silver, and Gold S3 layers, Glue PySpark transformations, dbt on Athena, and Redshift Serverless with Spectrum; orchestrated by MWAA, provisioned as modular Terraform across Dev, Staging, and Prod with GitHub Actions CI/CD.
- ▶ Built a natural language analytics agent on top of the lakehouse: users ask questions in plain English, the agent generates partition-aware Athena SQL, executes it, and returns plain-English insights with interactive charts. Stack: Claude API, FastAPI, Streamlit, ECS Fargate, with SQL guardrails, cost tracking, multi-turn sessions, and a full CI/CD deploy pipeline.
- ▶ Engineered high-throughput Kafka pipelines on Confluent Cloud streaming enriched sensor events into Snowflake, with raw data archived to S3 for replay and auditability.
- ▶ Delivered batch and incremental ELT pipelines with dbt across Snowflake, BigQuery, and Redshift for real estate, healthcare, and fraud detection use cases.
- ▶ Orchestrated multi-step workflows in Airflow, coordinating ingestion, transformation, data quality testing, and downstream analytics delivery.

Data Engineer

2022 to 2024

Digital Spine GmbH Berlin, Germany

- ▶ Built a real-time IoT data platform using Databricks Asset Bundles and Amazon Kinesis, processing over one million elevator telemetry records daily with sub-minute latency.
- ▶ Established GitHub Actions CI/CD pipelines for Databricks and dbt, ensuring consistent and reproducible deployments across all environments.
- ▶ Enforced data quality using PySpark and dbt: unit tests, integration tests, and quarantine tables to isolate invalid records before production ingestion.
- ▶ Built Power BI dashboards tracking elevator health metrics, downtime patterns, and predictive maintenance indicators for operational decisions.
- ▶ Communicated pipeline findings clearly to engineering, operations, and business stakeholders.

Data Specialist

2020 to 2022

Potters Real Estate Limited Nigeria

- ▶ Built ELT workflows using dbt and AWS (S3 and Redshift) to centralise and standardise real estate data, replacing manual processes with reliable automated pipelines.
- ▶ Analysed platform and marketing data to surface property demand trends, informing acquisition and marketing strategy.
- ▶ Deployed Power BI dashboards contributing to a 20 percent increase in lead generation.

PORTFOLIO

github.com/ChuquEmeka AWS medallion lakehouse, Databricks pipelines, Terraform modules, dbt and Airflow

[youtube.com/@Data Pipeline Lab](https://www.youtube.com/@DataPipelineLab) Technical walkthroughs of production data platform builds

CERTIFICATIONS

- ▶ Advanced Data Engineering with Databricks
- ▶ Apache Spark SQL for Data Analysts, Databricks
- ▶ Business Value with Data and Looker Specialization, Google
- ▶ Diploma in Data Engineering