

# Luke Agbetu

Aubrey, TX 76227

(989) 817-5630 | [luke.agbetu@gmail.com](mailto:luke.agbetu@gmail.com)

## PROFESSIONAL SUMMARY

---

Data Engineer with strong enterprise experience designing, building, and operating scalable data platforms across U.S. and international environments. Specialized in SQL-heavy data engineering, Spark-based processing, and production ETL/ELT pipelines using Python and Apache Airflow. Proven ability to own data pipelines end to end, improve data quality and reliability, and support analytics and reporting at scale across AWS, Azure, and GCP environments. Hands-on contributor in regulated and high-volume enterprise systems.

## CORE COMPETENCIES

---

- Enterprise Data Engineering & Platform Development
- Advanced SQL & Query Optimization
- ETL / ELT Pipeline Design & Automation
- Distributed Data Processing (Batch & Streaming)
- Workflow Orchestration (Apache Airflow)
- Dimensional Data Modeling (Star & Snowflake)
- Cloud Data Platforms (AWS, Azure, GCP)
- CI/CD & Infrastructure Automation
- Data Quality, Reliability & Observability
- Security & Compliance in Regulated Environments

## TECHNICAL SKILLS

---

### Databases & Data Warehousing

Oracle Database, PostgreSQL, MySQL, SQL Server, Snowflake, Amazon Redshift, Google BigQuery, MongoDB, Teradata, DB2

### Programming & Scripting

Python, SQL, PL/SQL, PowerShell, Java, R, C# (.NET Framework / .NET Core / ASP.NET), PHP

### Data Engineering & Big Data

Apache Airflow, Apache Spark (Batch & Streaming), Spark SQL, PySpark, Kafka, Flink, Hadoop (HDFS, Hive, Pig, Sqoop), dbt, Pandas, NumPy, ETL / ELT Pipelines

### Cloud Platforms & DevOps

AWS (S3, EC2, Redshift, Glue, EKS, CloudWatch)  
Azure (Data Factory, Databricks, Blob Storage, Azure SQL, HDInsight)  
GCP (BigQuery, Cloud Storage, Pub/Sub, Dataflow, Dataproc, Cloud Composer)  
Databricks (Spark, PySpark)  
Terraform, Docker, Kubernetes, GitHub Actions, CI/CD  
Microsoft Sentinel (SIEM), Azure Log Analytics, KQL

### Data Modeling & Architecture

Dimensional Modeling (Star & Snowflake Schemas), SCD (Slowly Changing Dimensions), Transactional Modeling, Medallion Architecture (Bronze / Silver / Gold)

### BI & Analytics

Power BI, Tableau, Oracle BI, Excel (Advanced Formulas, Pivot Tables), Matplotlib, Seaborn

### Machine Learning & Analytics

Scikit-learn, TensorFlow, Keras, Spark MLlib, Regression (Linear & Logistic), Decision Trees, Random Forests, Clustering, Feature Engineering

### Security, Monitoring & Governance

Rapid7 InsightVM, SolarWinds, Datadog, Vulnerability Management, Compliance, PII Protection

## Collaboration & Enterprise Tools

Git, GitHub, Jira, Confluence, Informatica, SSIS, SSRS, SSAS, IBM DataStage

## Operating Systems

Linux, Unix, Windows

## RELEVANT EXPERIENCE

---

### Data Engineer

AT&T | January 2025 – Present

- Designed, built, and owned production-grade ETL/ELT pipelines using Python, SQL, Spark, and Apache Airflow to process large operational and customer datasets.
- Developed batch and near-real-time Spark pipelines (PySpark, Spark SQL) optimized for distributed processing and scalability.
- Implemented data quality checks, schema validation, and anomaly detection, reducing data issues and reprocessing incidents.
- Modeled analytical datasets using star and snowflake schemas to support BI and downstream analytics use cases.
- Worked across AWS and Azure to manage ingestion, orchestration, and storage of analytical datasets.
- Supported CI/CD workflows using GitHub Actions, Docker, and Terraform for automated deployment of data pipelines.
- Worked with GCP services including BigQuery and Cloud Storage to support analytics use cases and cross-cloud data exploration.
- Used Databricks (Spark, PySpark) for distributed data processing and performance optimization of analytical workloads.
- Applied secure data handling practices to meet compliance and sensitive data protection requirements.
- Used Terraform to provision and manage cloud infrastructure supporting data pipelines, and worked with Microsoft Sentinel to ingest, analyze, and monitor security and operational logs.

### Tools & Technologies:

Python (Pandas, NumPy, Matplotlib), SQL, Power BI, AWS (S3, Glue, Redshift), Azure Data Factory, Docker, Kubernetes, Terraform, Apache Airflow, Snowflake, dbt, Excel, Git

### Data Engineer | Cloud & Analytics Support

Exact Trust, Dallas, TX. | October 2023 – December 2024

- Engineered and operated cloud-based data pipelines across AWS and Azure supporting analytics, reporting, and operational workloads.
- Built and orchestrated ETL pipelines using Spark, Python, SQL, and Apache Airflow for structured and semi-structured data.
- Integrated vulnerability management data (Rapid7) into analytics workflows for security monitoring and compliance reporting.
- Automated ingestion, transformation, and reporting workflows using Python and PowerShell, reducing manual operational effort.
- Implemented monitoring and observability using SolarWinds and log analytics to improve incident detection and response.
- Supported analytics and data processing workloads using Databricks and Spark across cloud environments.
- Authored SOPs, runbooks, and technical documentation for engineering and operations teams.

**Tools & Technologies:** AWS, Azure, Terraform, GitHub Actions, Python, PowerShell, Docker, Kubernetes, SQL.

## **Data Engineer / SQL Database Developer**

**DbConn** | Hamburg, Germany | March 2023 – September 2023

- Developed and optimized complex T-SQL stored procedures, views, and table-valued functions for enterprise SQL Server environments.
- Refactored legacy SQL workloads using indexing and execution-plan analysis to improve performance and scalability.
- Supported data extraction, transformation, and reporting workflows for analytics and downstream applications.
- Collaborated with application and reporting teams to translate business requirements into efficient database solutions.

## **Earlier Experience (Lead, Software Developer/Technology)**

**Skye Bank** | **UBA** | **Guaranty Trust Bank**, Sierra Leone, West Africa | 2015 – 2022

- Led enterprise data platforms, reporting systems, and governance initiatives in regulated financial environments.
- Supported digital banking analytics, regulatory reporting, data security, and platform modernization initiatives.

## **EDUCATION**

---

MSc. Computer Science – Towson University, Maryland, United States

December 2025

BSc. Computer Science & Mathematics - Fountain University, Osun State, Nigeria

July 2011

## **CERTIFICATIONS**

---

- CompTIA Security+ ce Certification,
- AWS Certified Cloud Practitioner Certification
- Certificate of Completion (R Certification)

## **PUBLICATION**

---

*Detecting SARS-CoV-2 in CT scans Using Vision Transformer & GNN* — MDPI Algorithms