

Luke Agbetu

Aubrey, TX 76227 | (989) 817-5630 | luke.agbetu@gmail.com | Portfolio: lanre2907.github.io

PROFESSIONAL SUMMARY

Lead Data Engineer with 9+ years of experience designing, building, and optimizing scalable data pipelines, cloud-native platforms, and enterprise data solutions across telecom, financial services, and regulated healthcare environments. Deep expertise in Python, SQL, PySpark, Spark, Databricks, and Apache Airflow, with production experience across AWS, Azure, and GCP. Proven record of leading end-to-end data engineering initiatives, from architecture and pipeline design through performance tuning, quality assurance, and stakeholder delivery. Skilled in data lake and warehouse architecture, distributed processing, API integration, schema design, ML pipeline support, and building reliable, observable data systems at scale.

CORE COMPETENCIES

- Enterprise Data Engineering & Platform Development
- Advanced SQL Engineering & Query Optimization
- ETL / ELT Pipeline Design & Automation
- Distributed Data Processing (Batch & Near Real-Time)
- Data Lakehouse & Warehousing Architecture
- Workflow Orchestration (Apache Airflow)
- Cloud Data Engineering (AWS, Azure & GCP)
- Dimensional Data Modeling (Star & Snowflake)
- PostgreSQL / SQL Server / Oracle Database Engineering
- Query Performance Tuning & Execution Plan Analysis
- Production Support, Runbooks & Operational Troubleshooting
- Data Quality, Reliability & Observability
- Security Automation (Microsoft Sentinel, Azure Logic Apps, Tines)
- Cross-Functional Stakeholder Collaboration
- Agile Delivery in Regulated Environments

TECHNICAL SKILLS

Languages & Core Tools: Python, SQL, PySpark, Spark SQL, PL/SQL, PowerShell, Java, Node.js, Express.js, C# (ASP.NET / .NET Core)

Data Engineering & Processing: Apache Spark (Batch & Streaming), Apache Airflow, Databricks, Delta Lake, Kafka, Apache Flink, Hadoop (HDFS, Hive, Pig, Sqoop), dbt, Pandas, NumPy, ETL / ELT Pipelines

Cloud Platforms: AWS (S3, Redshift, Glue, EKS, CloudWatch) | Azure (Data Factory, Databricks, Blob Storage, Azure SQL, HDInsight, Log Analytics, KQL) | GCP (BigQuery, Cloud Storage, Pub/Sub, Dataflow, Dataproc, Cloud Composer)

Databases & Warehousing: Snowflake, Amazon Redshift, Google BigQuery, Oracle Database, PostgreSQL, SQL Server, MongoDB, Teradata, DB2

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

DevOps & Infrastructure: Docker, Kubernetes, Terraform, GitHub Actions, CI/CD, Linux

Security & Automation: Microsoft Sentinel (SIEM), Azure Logic Apps, Tines (SOAR), Rapid7 InsightVM, SolarWinds, Vulnerability Management, Compliance, PII Protection

BI & Visualization: Crystal Reports, Power BI, Tableau, Excel (Advanced)

Monitoring & ML: Datadog, Scikit-learn, TensorFlow, Keras, Spark MLlib

Collaboration & Enterprise Tools: Git, GitHub, Jira, Confluence, Informatica, SSIS, IBM DataStage

RELEVANT EXPERIENCE

Data Engineer

January 2025 – Present

AT&T | Telecom

- Designed, built, and owned production-grade ETL/ELT pipelines using Python, SQL, Spark, and Apache Airflow, processing 10M+ daily customer and operational records across AT&T's telecom data platform.

- Developed batch and near-real-time Spark pipelines using PySpark and Spark SQL, reducing average processing latency by ~30% through partition tuning, broadcast joins, and execution-plan optimization.
- Implemented data quality checks, schema validation, and anomaly detection across 15+ pipeline DAGs, cutting data reprocessing incidents by ~40% and improving downstream reporting reliability.
- Modeled analytical datasets using star and snowflake schemas, supporting 8+ downstream BI and analytics teams consuming data through Redshift and Snowflake.
- Built CI/CD workflows using GitHub Actions, Docker, and Terraform, reducing pipeline deployment time from hours to under 15 minutes and enabling more reliable production releases.
- Optimized Databricks Delta Lake tables using Z-ordering and VACUUM, reducing query scan times by ~25% on high-volume workloads.
- Containerized 20+ data pipelines using Docker and deployed on Kubernetes (EKS), enabling auto-scaling and fault-tolerant execution across distributed workloads.
- Collaborated with product managers, analysts, solutions architects, and business stakeholders to translate complex reporting and data requirements into production-ready pipelines, models, and analytics-ready datasets.

Tools & Technologies: Python, SQL, PySpark, Spark SQL, Apache Airflow, Databricks, Delta Lake, Amazon Redshift, Snowflake, AWS (S3, EKS), Docker, Kubernetes, Terraform, GitHub Actions, Power BI, Git

Data Engineer | Cloud & Analytics Support

October 2023 – December 2024

Exact Trust | Dallas, TX

- Engineered and operated cloud-based data pipelines across AWS and Azure, supporting analytics and reporting for 6+ internal business units and processing 2TB+ of structured and semi-structured data monthly.
- Built Databricks and Spark-based Medallion pipelines using Delta Lake, reducing data processing time by ~35% and improving data freshness for downstream consumers.
- Designed and deployed security automation workflows using Microsoft Sentinel (SIEM) and Azure Logic Apps, automating threat detection, alert triage, and incident response to reduce manual analyst workload.
- Built end-to-end security orchestration playbooks using Tines (SOAR), automating multi-step response actions across Sentinel, email, and ticketing systems — cutting mean time to respond to incidents.
- Integrated Rapid7 vulnerability data into analytics dashboards and compliance workflows, improving incident detection response time by ~20%.
- Automated ingestion, transformation, and reporting workflows using Python and PowerShell, eliminating ~15 hours/week of manual effort across the team.

Tools & Technologies: Python, PowerShell, Apache Spark, Databricks, Delta Lake, AWS, Azure Data Factory, Microsoft Sentinel, Azure Logic Apps, Tines, Rapid7 InsightVM, SolarWinds, Power BI

Data Engineer / SQL Database Developer

March 2023 – September 2023

DbConn | Hamburg, Germany (Hybrid)

- Developed and optimized 30+ complex SQL objects, including stored procedures, views, and table-valued functions for enterprise database environments supporting 5+ business applications.
- Performed advanced query tuning, indexing strategy improvements, and execution-plan analysis, improving query performance by up to 60% on high-frequency reporting and transactional workloads.
- Built backend automation solutions using Python and SQL to streamline customer processes and improve operational efficiency.

Tools & Technologies: SQL Server, T-SQL, Crystal Reports, Python, Stored Procedures, Views, Indexing, Execution Plan Analysis, Git

Lead, Software Developer / Technology

2012 – 2022

Skye Bank | UBA | Guaranty Trust Bank | Unified Payment Services | Powersoft Integrated Solutions

- Led enterprise technology, digital banking, and data-driven platforms across regulated financial institutions in Nigeria and Sierra Leone, spanning application delivery, core banking integrations, database systems, and infrastructure modernization.
- Designed and implemented API-driven integrations using RESTful and SOAP services to connect core banking systems with mobile banking, internet banking, USSD, payment, transaction notification, and real-time reporting platforms.
- Delivered production systems including mobile account opening, real-time payment portals, SMS alert platforms, internet banking, USSD banking, POS/ATM settlement, and government-facing transaction reporting solutions.
- Led Oracle core banking upgrade initiatives and supported enterprise database environments, including disaster recovery, data center migration, and business continuity projects.
- Built and supported bespoke web and database applications using Oracle, SQL Server, PostgreSQL, MySQL, C#, Java, and .NET technologies, with hands-on experience in backend development and reporting.

- Developed and maintained business and operational reports using Crystal Reports and SQL-based reporting tools, supporting finance, transaction monitoring, and internal business workflows.

Tools & Technologies: Oracle, PostgreSQL, SQL Server, MySQL, PL/SQL, Crystal Reports, Java, C#, ASP.NET, .NET, Node.js, REST APIs, SOAP, Power BI, Git

EDUCATION

MSc. Computer Science

December 2025 (Completed)

Towson University, Maryland, United States

BSc. Computer Science & Mathematics

July 2011

Fountain University, Osun State, Nigeria

CERTIFICATIONS

- AWS Certified Data Engineer – Associate
- AWS Certified Cloud Practitioner
- CompTIA Security+ CE Certification

PUBLICATION

Detecting SARS-CoV-2 in CT Scans Using Vision Transformer & GNN — [MDPI Algorithms](#)