Luke Agbetu

Senior Data Engineer

Aubrey, TX 76227 | (989) 817-5630 | luke.agbetu@gmail.com | Portfolio: lukeagbetu.dev

PROFESSIONAL SUMMARY

Senior Data Engineer with over a decade of experience architecting and delivering enterprise data platforms across financial services, telecom, and regulated environments, including years leading technology delivery for banks and payment providers. Specializes in Azure and Databricks lakehouse engineering, scalable batch and incremental pipelines, dimensional and semantic modeling for Power BI, and metadata-driven data quality and governance, with production depth across AWS, Azure, and GCP. Owns work end to end: leading technical spikes and logical design, making architecture trade-offs, and taking solutions through Dev, UAT, Preprod, and Production in partnership with business, governance, and platform teams.

SELECTED IMPACT

- **Re-architected a core marketing data load at Fisher Investments,** moving history off on-premises systems to Azure and scoping the daily ADF refresh from full history to a rolling 2-year window.
- **Designed and presented an enterprise data quality framework** to technical and leadership stakeholders: metadata-driven Databricks checks for AUM data scoring 4 quality dimensions, governed by a DAMA-DMBOK2-aligned framework.
- **Own data delivery for 65+ marketing data assets** across 5 environments (Dev, UAT, Preprod, Production, on-premises) at Fisher Investments.
- **Cut pipeline latency by ~30% and data reprocessing incidents by ~40%** on a telecom platform processing **10M+ daily** records at AT&T.
- **Reduced deployment time from hours to under 15 minutes** by building CI/CD with GitHub Actions, Docker, and Terraform.
- **Improved query performance by up to 60%** and **saved ~15 hours/week** of manual effort through SQL tuning and workflow automation.

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 (Airflow, ADF)
- Cloud Data Engineering (AWS, Azure & GCP)
- Dimensional Data Modeling (Star & Snowflake)
- Semantic & Tabular Modeling for Power BI (DAX)
- PostgreSQL / SQL Server / Oracle Database Engineering
- Query Performance Tuning & Execution Plan Analysis
- Production Support & Runbooks
- Data Quality, Governance & Observability
- Technical Spikes & Solution Design
- Security Automation (Sentinel, Tines)
- Cross-Functional Stakeholder Collaboration
- Agile Delivery in Regulated Environments

TECHNICAL SKILLS

Languages: Python, SQL, T-SQL, PySpark, Spark SQL, PL/SQL, DAX, PowerShell, C# / .NET, Java

Data Engineering & Processing: Apache Spark (Batch & Streaming), Apache Airflow, Azure Data Factory, Databricks, Delta Lake, Kafka, dbt, Pandas, ETL / ELT Pipelines

Databricks Platform: Lakehouse, Delta Lake, Databricks Asset Bundles (DAB), Databricks Apps, Databricks SQL Dashboards

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

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

Data Modeling & Architecture: Star & Snowflake Schemas, SCD, Medallion Architecture (Bronze / Silver / Gold), Tabular / Semantic Models

Data Quality & Governance: Metadata-Driven DQ Frameworks, DQ Scorecards (Completeness, Timeliness, Uniqueness, Validity), Data Lineage, Metadata Management, Compliance Dashboards, DAMA-DMBOK2

BI & Semantic Layer: Power BI, Tabular Models, Tabular Editor, Visual Studio, Tableau

DevOps & Tooling: Git, GitHub Actions, CI/CD, Docker, Kubernetes, Terraform, Linux, Jira, Confluence

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

PROFESSIONAL EXPERIENCE

Senior Data Engineer June 2026 – Present

Fisher Investments | Plano, TX (Hybrid)

- **Own end-to-end engineering for the Marketing Data Mart** on Azure Data Factory, Databricks, Azure SQL, and SQL Server, delivering governed data for 65+ marketing data assets across Dev, UAT, Preprod, Production, and on-premises environments.
- **Architected the redesign of a core marketing dataset's load strategy:** authored the logical-design spike, re-platformed all-time history from on-premises systems to an Azure source, and refactored the ADF pipeline into a rolling two-year daily refresh, shrinking each daily run from full history to a 2-year window and removing on-prem reloads and redundant historical staging.
- **Led a cross-model key remediation,** moving Campaign and Responses facts and dependent dimension/fact pipelines to a composite primary key plus source-system identifier, with coordinated Tabular model changes, to eliminate cross-source key collisions and protect reporting integrity.
- **Architected a metadata-driven data quality framework** in Databricks for AUM data, scoring completeness, timeliness, uniqueness, and validity through an automated DQ scoreboard, and authored the DQ Governance Framework aligned to DAMA-DMBOK2.
- **Delivered performance-versus-goal analytics** by integrating campaign forecast targets (yearly cost-per-set and set-rate goals) into the Campaign and Responses facts and the Power BI semantic layer.
- **Lead technical spikes that shape the platform roadmap,** including Day-over-Day snapshot metrics and Power BI subscriptions, and reverse-engineering legacy report SQL to define requirements for its rebuild on the modern SQL Server / Power BI platform.
- **Presented the metadata-driven data quality framework design to technical and leadership stakeholders,** and author the design spikes and implementation plans the Marketing Data Mart team delivers from, setting the approach for history loads, key strategy, and snapshot reporting.
- **Build enterprise governance and compliance dashboards** in Databricks, and shipped a data observability Databricks App through Databricks Asset Bundles with Git-based, multi-environment deployment.
- **Engineer Tabular semantic models and DAX measures** (Visual Studio, Tabular Editor) with full and incremental processing across Dev/UAT/Preprod/Prod, giving the business one governed layer for Power BI reporting.
- **Partner with Data Governance, external implementation partners, and architects** to establish Databricks metadata management, lineage, and data standards, and translate business requirements into production-ready data products.

Tools & Technologies: Azure Data Factory (ADF), Databricks, Databricks Asset Bundles, Databricks Apps, Azure SQL, SQL Server, T-SQL, Python, Power BI, Tabular Models, DAX, Tabular Editor, Visual Studio, Data Governance, Data Quality, Dimensional Modeling, Lakehouse Architecture, Jira, Git

Data Engineer January 2025 – May 2026

AT&T | Dallas, TX (Remote)

- 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 (Remote)

- 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 security automation with Microsoft Sentinel, Azure Logic Apps, and Tines (SOAR) playbooks, automating threat detection, triage, and multi-system incident response to cut analyst workload and response times.
- 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.
- Improved query performance by **up to 60%** on high-frequency reporting and transactional workloads through indexing strategy, query tuning, and execution-plan analysis, and built Python/SQL automation for customer processes.

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

Technology Lead / Lead Software Developer 2012 – 2022

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

- **Led enterprise technology and digital banking delivery** across regulated financial institutions in Nigeria and Sierra Leone, spanning core banking integration, database platforms, and infrastructure modernization.
- **Architected REST and SOAP integrations** connecting core banking with mobile, internet, USSD, payment, notification, and real-time reporting channels, and delivered POS/ATM settlement and government-facing transaction reporting systems.
- **Led Oracle core banking upgrades** and disaster recovery, data center migration, and business continuity programs.
- **Built database applications and operational reporting** on Oracle, SQL Server, PostgreSQL, and MySQL for finance and transaction monitoring.

Tools & Technologies: Oracle, PL/SQL, SQL Server, PostgreSQL, MySQL, C#, .NET, Java, REST / SOAP APIs, Crystal Reports

EDUCATION

MSc. Computer Science and Information System December 2025 (Completed)

Towson University, Maryland, United States

CERTIFICATIONS

- Databricks Certified Data Engineer Associate (2026)
- 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