

Kalu Okonkwo

Senior Data Engineer

Duluth, GA | ksonnie1@gmail.com | www.kaluokonkwo.com | github.com/kkalu-stack

SUMMARY

Senior Data Engineer with 8+ years across data engineering and analytics, building cloud data platforms for Ford, Amgen and Bayer. Specializes in Azure, Databricks, Snowflake, PySpark and dbt, with pipelines processing millions of manufacturing records daily and a Snowflake platform that consolidated 12 source systems into one enterprise data model. Led a five-person engineering and data science team and delivered data systems for regulated pharmaceutical operations. Holds a Doctor of Computer Science and builds production data products with LLM integration and agentic AI.

TECHNICAL SKILLS

Cloud: Azure (ADF, Synapse, ADLS, Azure DevOps), AWS, GCP

Data Platforms: Databricks, Snowflake, Delta Lake, Apache Spark, medallion (Lakehouse) architecture

Engineering: Python, PySpark, SQL, R, Pandas, NumPy, dbt, Airflow, SSIS, data modeling, data warehousing

Databases: PostgreSQL, SQL Server, Oracle, MySQL, Azure SQL

DevOps: Git, Docker, Azure DevOps CI/CD, GitHub Actions, SonarQube, pytest, JIRA, Agile/Scrum

AI and BI: Agentic AI, Anthropic and OpenAI APIs, Scikit-learn, Tableau, Power BI, Qlik Sense, Spotfire

Governance: FDA/GxP, HIPAA, GDPR, FedRAMP, data quality frameworks

EXPERIENCE

Senior Data Consultant | Spintiq (independent consulting), Remote **Apr 2025 to Present**

- Designed and built a nightly batch data pipeline in Python that ingests market data after every close, runs validation checks at ingestion, computes point-in-time features and writes ranked output for a stock analytics platform.
- Engineered a walk-forward backtesting framework with point-in-time controls to prevent lookahead bias, and a modular interface that lets data providers change without modifying downstream logic.
- Built an autonomous AI agent on Cloudflare Workers, Durable Objects and D1 that scans live markets every few minutes, scores opportunities with predictive models, sizes positions, places and exits orders through an exchange API, and reconciles fills and P&L, logging every decision for audit.
- Integrated the Anthropic and OpenAI APIs into production applications, turning structured data into plain-language analysis with streaming responses.
- Designed the Postgres data model and storage layer for a subscription platform on Supabase and Cloudflare R2, with signed-URL access control, multi-currency payments and users in 8 countries.

Python, TypeScript, Cloudflare Workers, D1, PostgreSQL (Supabase), Cloudflare R2, Anthropic API, OpenAI API, pytest

Data Engineer (Contract) | Amgen, Remote **Apr 2024 to Jun 2025**

- Built automated PySpark pipelines on Databricks that reduced drug product data processing time by **40%**, getting data to manufacturing, supply chain and regulatory teams faster.
- Implemented a layered medallion data model on Delta Lake across ingestion, staging and curated zones, giving downstream teams consistent, query-ready data.
- Led UAT across 3 cross-functional teams for the launch of Drug Product Data Systems, validating 200+ fields against FDA/GxP requirements with zero defects at launch.
- Wrote rule-driven validation logic and automated checks on stability testing data that caught formulation drift early and lowered the risk of production failures.
- Designed role-based access controls and encryption for sensitive batch record systems supporting annual product reviews.

Databricks, Delta Lake, PySpark, SQL, dbt, ADF, Azure

Lead Data Engineer (Contract) | Bayer, Remote **Aug 2023 to Dec 2023**

- Led a team of 5 data engineers and scientists to build a Snowflake analytics platform that consolidated 12 siloed sources into one enterprise data model, which became the single source of truth.
- Redesigned global KPI reporting in Tableau and Power BI across 4 business units, cutting executive report generation time by **80%**, from days to hours.
- Built SQL and Python workflows for large-scale data profiling and transformation, and set enterprise governance standards that improved data quality and auditability.

Snowflake, Azure, Python, SQL, Tableau, Power BI

Lead Data Engineer | Ford Motor Company, Remote

Aug 2021 to Jun 2023

- Designed end-to-end Azure data pipelines with PySpark, Databricks and dbt that cut data latency by **30%** while processing millions of manufacturing records a day across plant environments.
- Built automated ingestion and transformation frameworks handling TB-scale data that became the backbone of enterprise analytics and reporting across global operations.
- Deployed Tableau and Power BI dashboards used daily by **50+ managers** across manufacturing and logistics.
- Integrated SonarQube into Azure DevOps CI/CD pipelines, improving pipeline reliability and reducing production incidents and technical debt.

Azure, Databricks, PySpark, dbt, Azure DevOps, SonarQube

Brightly Software | Grand Junction, CO

Senior Data Analyst

Feb 2020 to Aug 2021

- Built Python and SQL predictive models that forecast customer energy consumption patterns and used them to target marketing campaigns, driving a **15%** increase in campaign conversion rates.
- Owned the ETL pipelines across SQL Server and Snowflake that fed executive reporting, consistently meeting data freshness SLAs.
- Delivered Tableau dashboards on energy utilization KPIs that executives used to make cost-saving decisions.
- Mentored junior analysts in data modeling, visualization and reporting, shortening their ramp-up time and raising output quality.

Data Analyst

Sep 2018 to Feb 2020

- Wrote complex SQL to clean, transform and integrate data from multiple source systems into a single source of truth for company reporting.
- Automated recurring reporting with Python workflows, cutting manual reporting time by **30%**.
- Built automated Tableau dashboards tracking product usage, customer churn and retention for cross-functional stakeholders.

Python, SQL, SQL Server, Snowflake, Tableau

EDUCATION

Doctor of Computer Science Colorado Technical University	2026
M.S. Data Analytics Maryville University	2019
B.S. Petroleum Engineering University of Wyoming	2017

CERTIFICATES

Applied Data Science Certificate MIT Professional Education	2023
Product Management Certificate Northwestern University	2021

RESEARCH

FILM-DTI, independent research manuscript (sole author). Graph neural network for drug-target interaction prediction, evaluated on 425,845 BindingDB pairs.