

Ayotomiwa Salau

Senior Senior Data Engineer

SUMMARY

- Senior Data Engineer with 8+ years in fintech, banking, and telecom, skilled in Python, Java, SQL, and cloud-native tools like Airflow, Spark, Kafka, Redshift, and Snowflake.
- Expertise in building scalable ETL pipelines, real-time data processing, and AI integrations using Huggingface and Kubernetes, improving data integrity and reducing pipeline breakage by 85%.
- Proficient in cloud infrastructure automation with Terraform, CloudFormation, and container orchestration using Docker and Kubernetes.
- Strong background in Agile methodologies (Scrum, Kanban), TDD, CI/CD pipelines, and delivering high-impact data engineering solutions across multiple international organizations.

TECHNICAL SKILLS

Main Technical Skills	Python (8 yr.), Java (7 yr.), SQL (7 yr.), ETL (7 yr.), Apache Spark (6 yr.)
Programming Languages	Java (7 yr.), Python (8 yr.), VBA (5 yr.)
Java Frameworks	Apache Spark (6 yr.), Spring Boot (6 yr.)
Scala Frameworks	Apache Spark (6 yr.)
AI & Machine Learning	Hugging Face (4 yr.), PandasAI (7 yr.)
.NET Platform	VBA (5 yr.)
Data Analysis and Visualization Technologies	Apache Airflow (6 yr.), Apache Spark (6 yr.), ARIMA (5 yr.), AWS Athena (5 yr.), Azure Data Factory (4 yr.), Data modeling Kimbal (5 yr.), ETL (7 yr.), Looker Studio (4 yr.), PandasAI (7 yr.), Power BI (5 yr.), Tableau (5 yr.)
Databases & Management Systems / ORM	Apache Spark (6 yr.), AWS Redshift (6 yr.), Data Warehousing (6 yr.), Microsoft SQL Server (5 yr.), Oracle Database (6 yr.), PostgreSQL (7 yr.), Snowflake (5 yr.), SQL (7 yr.), T-SQL (5 yr.)
Cloud Platforms, Services & Computing	GCP (6 yr.)
Amazon Web Services	AWS Athena (5 yr.), AWS Cloudformation (5 yr.), AWS EMR (5 yr.), AWS Glue (5 yr.), AWS Redshift (6 yr.), AWS S3 (6 yr.)
Azure Cloud Services	Azure Data Factory (4 yr.)
Google Cloud Platform	Google BigQuery (5 yr.)
Deployment, CI/CD & Administration	CI/CD (6 yr.)

Virtualization, Containers and Orchestration	Docker (6 yr.), Kubernetes (6 yr.), Terraform (5 yr.)
QA, Test Automation, Security	JUnit (5 yr.)
Message/Queue/Task Brokers	Kafka (5 yr.)
Methodologies, Paradigms and Patterns	Kanban (6 yr.), Scrum (6 yr.), TDD (7 yr.)
SDK / API and Integrations	Keycloak (5 yr.)
Other Technical Skills	Metabase (4 yr.)

WORK EXPERIENCE

<https://ayotomiwasalau.com>

Senior Data Engineer - Moniepoint Group (Data Engineering and Cloud Infrastructure)

Duration: November 2025 - Present

- Designed and maintained high-volume Airflow data pipelines for financial reconciliations and settlement processing, improving bulk data processing speed and helping the business consistently meet a 2-hour resolution SLA.
- Ingested financial and operational data from Google Cloud Spanner into ClickHouse using Kafka & Debezium for analytics, reporting, and transaction investigation, enabling faster access to reconciled settlement data.
- Built and deployed scalable batch processing jobs on Google Cloud Dataflow, processing large transaction datasets into sources such as BigQuery and reducing manual operational effort across reconciliation workflows.
- Analysed large datasets using Python and SQL to investigate erroneous transactions, identify data quality issues, and support reconciliation accuracy across high-volume payment records.
- Optimised ClickHouse data models, queries, and platform performance for speed, scalability, and cost efficiency, improving analytical query performance using indexing and partitioning by up to 40%.

Senior Data Engineer - Bravely [Andela] (AI Sentiment Pipeline and Data Quality Automation)

Duration: October 2023 - May 2025

Summary:

- Implemented AI-driven sentiment analysis and data quality automation pipelines to enhance customer feedback insights and data integrity for a US-based client
- Developed tools to manage model dependencies and improved survey data accessibility



Responsibilities:

- Implemented an AI sentiment pipeline using Airflow and Huggingface pretrained LLM model.
- Built Tableau dashboards for data visualization.
- Engineered Airflow pipelines for daily data quality tests on Redshift.
- Developed Python CLI tools for managing model dependencies in transformation jobs.
- Built Zoom survey ingestion and transformation data pipelines.

Technologies: Airflow, Huggingface LLM, Tableau, Python, n-trees, hashmaps, Pandas, Redshift, SQL

Senior Engineer - Andela (Cloud Native Data Engineering Projects)

Duration: September 2023 - May 2025

Summary: Delivered high-quality cloud native data engineering solutions leveraging best practices and modern data tools as an independent contractor placed with external clients.

Responsibilities:

- Delivered engineering outcomes using cloud native data tools and best practices.

Big Data Engineer - Symphony Solutions (Enterprise Observability and Marketing Data Ingestion)

Duration: December 2022 - October 2023

Summary: Developed an enterprise observability platform by ingesting and transforming Open Telemetry data and integrated marketing data ingestion pipelines to improve data processing efficiency and visualization.

Responsibilities:

- Set up Kafka Streams broker to ingest and transform Open Telemetry data into Druid timeseries DB.
- Integrated Prometheus and Grafana for metrics visualization.
- Built Java CRUD REST service with Spring Boot, KeyCloak, and PostgreSQL.
- Managed libraries with Gradle and schema evolution with Flyway.
- Established JUnit test suites for REST endpoints and Kafka Streams.
- Implemented streaming topic on GCP Pub-Sub for marketing data ingestion.
- Applied Kimball-style data modelling and developed Looker dashboards.

Technologies: Kafka Streams, Java, Spring Boot, KeyCloak, PostgreSQL, Gradle, Flyway, JUnit, Google Cloud Pub-Sub, DataProc, BigQuery, Kimball data modelling, Looker, Prometheus, Grafana, Open Telemetry, Druid

Lead Data Engineer - Indicina (Big Data Processing and Data Science Workflow Automation)

Duration: June 2021 - August 2023

Summary:

- Led big data processing and automated data science workflows to enhance transaction processing efficiency and regulatory compliance
- Built a data lake to improve cross-team data accessibility and self-service analytics



Responsibilities:

- Deployed Spark ETL jobs on AWS Hadoop EMR to process large transaction datasets.
- Automated data science workflows using Airflow, EMR, and Sagemaker.
- Led team to encrypt PII data in S3 ensuring GDPR compliance.
- Built data lake on S3 with ingestion, transformation, and analytics capabilities.

Technologies: Spark, AWS Hadoop EMR, Airflow, Docker, EKS Kubernetes, Sagemaker, AWS S3, AWS Glue, Lambda, Python, Athena (Presto), Metabase, AES encryption

Data Specialist - KPMG (Audit Data Solutions and Data Governance)

Duration: September 2018 - May 2020

Summary: Delivered audit data solutions for revenue assurance and risk projects, built ETL pipelines, automated data validation, and led data governance and privacy reviews to improve compliance and reduce audit risks.

Responsibilities:

- Built ETL pipelines with Azure Data Factory, T-SQL, and MSSQL Server.
- Ingested big data from Oracle databases using automated SQL stored procedures.
- Implemented automated data validation and assurance checks with IDea and VBA.
- Led data governance and privacy reviews assessing PII handling and internal controls.
- Designed remediation plans to improve data compliance.
- Built Statistical/ML applications for stock price modeling as proof of concept.

Technologies: Azure Data Factory, T-SQL, MSSQL Server, Oracle, SQL stored procedures, IDea, VBA, Power BI, Python, ARIMA model

EDUCATION

- **Udacity Nanodegree, Cloud Native Application Architecture**
August 2024
- **Udacity Nanodegree, Machine Learning DevOps Engineer**
September 2024
- **Udacity Nanodegree, Deep Learning**
September 2024
- **Udacity Nanodegree, Data Engineering**
September 2020
- **Udacity Nanodegree, Full Stack Developer**
July 2020

CERTIFICATION

- Introduction to Python for Data Science
- RPA Orchestrator 2018.3 Training - Level 2
- UiPath RPA Certified Advanced Developer
- Project Management Training Certification
- Analyzing and Visualizing Data with Excel

