

James Adeshina

Data Engineer - Cloud Pipelines | Python, SQL, AWS, BigQuery, Airflow | ETL/ELT & Data Modelling

(+44) 7780274687, jamesadeshina413@gmail.com, [LinkedIn](#), [Portfolio](#)

Results-driven Data Engineer with an MSc in Big Data Analytics (Distinction) and hands-on experience delivering automated, cloud-native data pipelines across AWS, Azure, and BigQuery. Specialising in Python, SQL, and modern ETL/ELT pipelines to transform complex, multi-source data into analytics, and quality datasets. Skilled in building structured warehouse models, enforcing data quality, lineage, and governance, and partnering with product, risk, and operational teams across public-sector, academic, and fintech environments.

PROFESSIONAL EXPERIENCE

University of Derby

AI/Data Engineer (*Research Project Associate*)

Derby, UK

03/2025 - Present

- Built an image-based assessment system using CNNs and OpenCV, with scalable ETL pipelines for ingesting, cleaning, and annotating large volumes of medical image data.
- Engineered end-to-end ML pipelines covering model training, evaluation, and real-time deployment for rehabilitation analysis.
- Developed an AI-powered Digital Rehabilitation Assistant (AI-DRA) for stroke survivors, leveraging computer vision and machine learning to deliver personalised therapy recommendations and support functional recovery.
- Designed and implemented a Functional Movement Screening (FMS) model using deep learning to analyse human movement patterns and quantify rehabilitation progress.
- Visualised FMS-derived analytics in Power BI, enabling physiotherapists to track progress across 20+ movement metrics, identify movement asymmetries, and tailor interventions with data-driven confidence.

University of Derby / Bolsover District Council

Data Engineer (NLP & Pipelines)

Derby, UK

01/2025 - 06/2025

- Led development of an NLP system automating analysis of 987 community consultation letters for a UK local authority, replacing manual triage with AI-driven classification, issue tagging, sentiment detection, and summarisation.
- Designed a multi-step pipeline to classify letters into objection and non-objection, extract key issues mapped to five grouped policy areas, detect sentiment (positive/negative/neutral), and generate summaries at three levels of detail (succinct, detailed, very detailed).
- Built a batch-processing interface with a 97% success rate (67/69) on stress testing, reducing manual review time from 6 to 10 hours to under one minute per 69 letter batch, a 95% efficiency gain.
- Created dashboards for department-level routing, sentiment trends, and issue distribution, enabling policy teams to visualise public concerns at scale and act on high-risk feedback faster.
- Demonstrated the feasibility of AI-driven correspondence analysis for public sector workflows by delivering structured outputs across five insight layers, enabling three department-level teams to triage feedback faster. Presented findings to 8 council board members, answered 5 live questions, and incorporated feedback into the system.

Amdari (Work Experience Program)

Data Engineer (Consulting Project)

Remote, United Kingdom

03/2023 - 10/2024

- Built an automated ETL pipeline for PrimeSquare Properties, processing 10,000+ property records monthly, with validation checks that improved downstream data quality.
- Migrated 5GB+ of on-prem PostgreSQL data to AWS S3 via Airflow, implementing a Bronze-Silver-Gold architecture with Parquet compression, reducing storage footprint ~20% and improving Athena query times.
- Set up automated schema validation that prevented ~15% ingestion errors during testing and ensured consistent data ingestion.
- Registered schemas with AWS Glue and enabled Athena query acceleration (3× faster), supporting revenue, pricing, and agent-performance reporting.

- Delivered Power BI dashboards that cut manual reporting workload by ~50%, while maintaining code quality through GitHub workflows and structured version control.

PROJECTS

Explainable AI for Diabetes and Pre-diabetes Detection (MSc Dissertation)

08/2025

Built a full ML training pipeline using Python and SQL, processing 100k+ hospital records and generating structured, analysis ready datasets. Developed scalable feature engineering, label encoding, and data validation routines. Trained Logistic Regression, Random Forest, and XGBoost models, applying SHAP for global and subgroup interpretability. Integrated GPT-4 to produce clear, human-readable explanations, linking data engineering workflows with clinical decision-support needs.

Monzo User Sentiment & Feature Insights Dashboard

10/2025

Engineered a Python-BigQuery data pipeline to ingest and process 33k+ app-store reviews with automated sentiment and keyword extraction. Published GDPR-compliant, enriched datasets to Power BI for trend analysis and product experience insights.

Retail Sales Integration Pipeline with AWS S3 and Airbyte

10/2025

Developed an automated ELT pipeline to centralise sales, customer, and product data for a retail company. Configured AWS S3 as a cloud data lake with versioning and structured hierarchy using Python (boto3). Integrated Airbyte to extract data from S3 and load it into a MotherDuck warehouse on a scheduled sync. Implemented schema validation, record-count checks, and SQL-based profiling for data accuracy and transparency across the pipeline.

TECHNICAL SKILLS

Programming & Query: Python (Pandas, PySpark), Scala (foundational knowledge), SQL (CTEs, optimisation), Java, REST APIs

ETL & Pipelines: ETL and ELT design, data ingestion, validation, orchestration (Airflow DAGs)

Cloud & Warehousing: AWS (S3, Redshift, Lambda, Glue), BigQuery, Azure Synapse

Data Modelling: Star and Snowflake schemas, dimensional modelling, data marts, Delta Lake

Spark & Lakehouse: PySpark, Delta Lake fundamentals, Databricks core workflows

IaC & Automation: Terraform (foundational), CI/CD, Airbyte, Power Automate, logging, monitoring, versioning

Databases: PostgreSQL, MySQL, SQL Server, Hive

Governance & Security: GDPR, data lineage, access controls, NHS Data Security Toolkit

Machine Learning (supporting): Scikit-learn, SHAP, feature engineering, model explainability for analytics use cases

ACHIEVEMENTS

Virgin Media O2

3/2025

Hackathon Winner (Insights Engineer Role)

Designed a data transformation pipeline for anonymised telecom datasets and optimised Power BI dashboards using DAX and SQL for sub-second performance. Recognised by product and engineering mentors for clean modelling, fast refresh logic, and analytical storytelling.

BLOC Digital

11/03/2025

AI Interface Design (Certificate of Achievement)

Recognised for contributing to industry-led prototypes involving data-driven interaction flows and model-to-UI integrations, demonstrating strong analytical and user-centred design thinking.

EDUCATION

UNIVERSITY OF DERBY

09/2024 - 11/2025

MSc Big Data Analytics

Modules included Processing Big Data (90%), Business Analytics (84%), and NLP (80%).

Dissertation: Explainable AI for Diabetes and Pre-diabetes Detection: Integrating SHAP and OpenAI GPT-4 for Human-Centred Explanations (84%). Built scalable pipelines and explainable ML workflows using Python, SQL, and SHAP.

BOWEN UNIVERSITY

09/2016 - 10/2020

BSc Computer Science & Information Technology (Second Class Upper, CGPA: 3.61/4.00)

Final-year project: AI-driven diagnostic chatbot for early-stage lung cancer detection.

CERTIFICATIONS

- AWS Certified Data Engineer Associate (DEA-C01) - LinkedIn Learning, 2025
- AWS Certified Cloud Practitioner (CLF-C02) - LinkedIn Learning, 2025
- Microsoft Azure Fundamentals (AZ-900) - LinkedIn Learning, 2025

- Complete Guide to Google BigQuery for Data and ML Engineers - LinkedIn Learning, 2025
- Apache PySpark by Example - LinkedIn Learning, 2025
- Docker for Data Engineers - LinkedIn Learning, 2025