

GEORGE IFI

Middlesbrough, UK · Mobile: +44 7490 199070 · Email: back2ifi@gmail.com
LinkedIn: linkedin.com/in/ifi-george-ifi · GitHub: github.com/REALJIFI

Professional Summary

Data Engineer with 4+ years delivering secure, scalable, and cost-efficient data platforms across AWS and Azure. Specialist in lakehouse architectures and distributed processing, Databricks, Apache Spark (PySpark, Structured Streaming), and Delta Lake, plus robust ETL/ELT design, Terraform-based Infrastructure-as-Code, and cost-aware warehousing (Snowflake / Redshift / BigQuery). Track record enabling self-service analytics with Athena/Power BI/Grafana and governed access (IAM/KMS, Unity Catalog, data quality).

Core Skills (ATS Keywords)

Lakehouse & Distributed Processing: Databricks (Azure & AWS), Delta Lake, Unity Catalog, Jobs/Workflows; Apache Spark (PySpark, Structured Streaming)

Cloud & IaC: Azure (ADLS Gen2, Data Factory, Synapse serverless SQL, Databricks, Key Vault, Private Endpoints, Monitor/Log Analytics); AWS (S3, Glue, Athena, IAM, KMS, Lake Formation); GCP; Terraform; Docker; CI/CD (GitHub Actions)

Orchestration & Streaming: Apache Airflow (DAGs), Event-driven pipelines; Kafka (foundational)

Data Warehousing: Snowflake, Amazon Redshift, Google BigQuery

Databases: SQL Server, PostgreSQL, MySQL

Programming: Python, SQL, Java

Data Management & Quality: Data modeling (star/snowflake), partitioning, Parquet/Delta, CDC, data quality checks, profiling, version control (Git)

BI & Monitoring: Power BI, Grafana, Excel

Professional Experience

Data Engineer Consultant — 10Alytics, Manchester, UK | Jan 2024 – Present

- Built scalable ETL/ELT pipelines with Apache Spark (PySpark) & Python, improving end-to-end processing time by 30% and accelerating reporting cycles.
- Delivered self-service analytics with Power BI and Grafana, reducing ad-hoc request turnaround and improving stakeholder satisfaction.

- Optimized Snowflake warehouses (clustering, caching, task scheduling), lowering compute/storage costs by ~£10k per quarter.
- Tuned SQL Server/PostgreSQL (indexes, query plans, partitioning) for 40% faster read/write performance in operational data stores.
- Implemented Azure Data Factory (ADF) ingestion landing to ADLS Gen2; triggered Azure Databricks PySpark jobs to write Delta tables; cut end-to-end latency by 35%.
- Hardened governance with Unity Catalog (Azure Databricks) and secret management via Azure Key Vault; streamlined access reviews and auditability.
- Enabled ad-hoc analytics via Synapse serverless SQL over Delta/Parquet in ADLS; reduced BI prep time by 40%.
- Secured data paths with Private Endpoints/VNET-integrated Databricks; eliminated public egress for sensitive datasets.

Data Engineer — Greater Women Initiative for Health & Rights (GWIHR), Nigeria | Jul 2019 – Jul 2023

- Automated ETL with Python, Airflow, and SSIS for program metrics; reduced manual handling by ~30% and improved data timeliness.
- Built data validation and profiling to detect outliers and schema drift; improved reliability of reporting datasets.
- Authored KPI frameworks and dashboards (Power BI/Grafana) that drove evidence-based decisions across public-health projects.
- Delivered SQL-based analytics and standardized data models to streamline weekly performance reviews and donor reporting.
- Coordinated Data Quality Assurance (DQA) and trained field teams on high-quality data capture and governance.

Assistant Monitoring & Evaluation Officer — GWIHR, Rivers State, Nigeria | Mar 2018 – Apr 2019

- Produced dashboards and visualisations; maintained KPI trackers and monthly performance reviews.
- Ensured timely and accurate data entry, documentation, and reporting across LGAs; supported capacity-building for community facilitators.
- Analyzed field data to present insights on achievements, gaps, and corrective actions.

Education

MSc Environmental Management (with Advanced Practice) — Teesside University, Middlesbrough, UK

B.Sc. Environmental Management — Chukwuemeka Odumegwu Ojukwu University, Anambra State, Nigeria

Selected Tools & Platforms

Databricks, Apache Spark (PySpark, Structured Streaming), Delta Lake, Airflow, Terraform, Docker, Azure (ADLS Gen2, Data Factory, Synapse serverless SQL, Databricks, Key Vault, Private Endpoints, Monitor/Log Analytics), AWS (S3, Glue, Athena, IAM, KMS, Lake Formation), Snowflake, Redshift, BigQuery, SQL Server, PostgreSQL, MySQL, Power BI, Grafana, Git