

# Bolaji Balogoun

## Senior Data Engineer | Cloud Data Platforms, Modeling & Reliability

Little Rock, AR, USA | hello@bolablg.com | 501-519-3150 | linkedin.com/in/bolablg | bolablg.com

### PROFESSIONAL SUMMARY

---

Senior Data Engineer with over 7 years building production data pipelines and platforms, backed by a decade in data science and advanced analytics. Built the cloud data foundation for a multinational startup as its first data hire and migrated tens of terabytes to BigQuery. Across progressive roles, delivered 250+ production pipelines, scaled daily upsert processing to approximately 5 TB across five markets, and reduced Google Cloud spend by over 42% while maintaining freshness and integrity SLAs.

### TECHNICAL SKILLS

---

**Cloud & Data Lakehouse:** GCP (BigQuery, GCS, BigLake, Dataplex, Dataform, Dataflow, Cloud Run, Pub/Sub), AWS (EC2, S3), Databricks (Delta Lake, Unity Catalog, Lakeflow)

**Data Engineering:** Python, SQL, PostgreSQL, MongoDB, Spark, batch and near-real-time pipelines, ETL/ELT, Airflow

**DataOps & Delivery:** Git, CI/CD, Docker, automated testing, code review, monitoring, operational support

**Modeling & Quality:** SCD & Fact modeling, Incremental modeling, partition-aware MERGE, assertions, monitoring, lineage, access controls

**Analytics & Communication:** Looker Studio, Metabase, Grafana, Google Workspace, metrics layers, executive reporting, stakeholder presentations

### PROFESSIONAL EXPERIENCE

---

**Gozem - Data Director** | Apr 2025 - Jul 2026 | Remote - five African countries and Singapore HQ

- Designed Airflow pipelines that transformed MongoDB, BigQuery, Google Cloud Storage, and Google Sheets data into application-ready datasets powering company-wide services on a platform serving 650+ employees and partners.
- Implemented 17 of the initial 19 microservices for a Python/FastAPI-based Data Hub in under one month; modeled schemas and persistence patterns for application state, workflows, and usage data.
- Developed a mobile-first FastAPI/Airflow application that onboarded roughly 700 external agents and reached about 200 daily users, with automated synchronization, retries, and operational monitoring.
- Operated a governed platform processing approximately 5 TB of daily upserts across five markets, spanning nearly 300 Dataform models and Airflow DAGs and sourcing data from 100+ MongoDB collections.
- Architected Airflow and BigQuery quality monitoring across hundreds of source and analytical tables, with alerts routed to owners and structured review workflows.
- Standardized Data Hub service reliability with health checks, Prometheus metrics, retries, targeted pytest suites, security scanning, and GitLab CI/CD.

**Gozem - Head of Data** | Oct 2022 - Mar 2025 | Cotonou, Benin - Hybrid

- Re-architected dependency-blind BigQuery scheduled queries into layered Dataform workflows with normalized source staging, incremental transformations, shared marts, partition-aware MERGE logic, targeted assertions, and rolling reprocessing for late-arriving events.
- Reduced Google Cloud spend 42.57% in one quarter by classifying usage across ingestion, transformation, reporting, and direct analysis, then using BigQuery job metadata to eliminate redundant execution and optimize expensive workloads.
- Deployed open-source Metabase and ClickHouse on AWS EC2, then used Airflow to replicate selected BigQuery tables and partitions with 3-to-18-month history windows for lower-cost analytical serving.
- Deployed self-hosted Airflow on AWS EC2 as the production orchestrator, replacing cron jobs; standardized DAG development with staging, dry runs, automated checks, code review, and GitLab CI/CD while onboarding a junior data engineer.
- Built a near-real-time MongoDB CDC pipeline that published change events to Pub/Sub, processed them with Dataflow, and updated BigQuery for e-commerce monitoring.

## **Gozem - Global Data Analyst** | Nov 2019 - Sep 2022 | Cotonou, Benin

- Automated 50+ Linux/cron ETL/ELT pipelines powering reporting, fraud screening, KPI alerts, and data sharing; migrated tens of terabytes from PostgreSQL and MongoDB to BigQuery within one quarter without operational interruption.
- Engineered ingestion from operational MongoDB sources and from Firebase/Google Analytics, Google Ads, and Meta Ads into BigQuery for product analytics, A/B test analysis, and campaign performance.
- Developed dimensional and incremental models for drivers, customers, merchants, trips, and transactions, including SCD Type 1 and Type 2 patterns and reusable metrics consumed by analytical and operational workflows.
- Designed multi-market models for acquisition, revenue attribution, cohorts, and unit economics across multiple currencies, including processing for late-arriving events.

## **SELECTED CONSULTING & EARLIER EXPERIENCE**

---

### **International Trade Centre (UN Agency) - Data Management Consultant** | Aug 2023 - Sep 2023 | ECOWAS

- Trained 15 ECOWAS startups for the UN ITC NTF V program; created reusable materials covering data collection, processing, cloud platforms, and operational applications, and coached five founders.

### **Rintio - Data Scientist** | Jul 2018 - May 2019 | Cotonou, Benin

- Built an on-premises Hadoop data lake and distributed architecture using Spark for processing and Elasticsearch for indexing, serving 100GB+ of agricultural data through Kibana and Apache Superset.

### **Rintio - Big Data Apprentice** | Mar 2018 - Jun 2018 | Ouidah, Benin

- Completed an intensive Data Engineering, Machine Learning, and Big Data program using Spark, Hive, Python, and SQL; contributed to the Duniya Data agricultural insights project.

### **INStaD Benin - Statistician** | Sep 2017 - Feb 2018 | Cotonou, Benin

- Structured public-sector survey data collection and validation workflows, consolidated field records, and delivered hands-on R training to approximately 30 agents across two departments.

### **SARL Le Pantagruel - Statistician** | Aug 2013 - Oct 2013 | Abomey-Calavi, Benin

- Designed the survey questionnaire and collection process, gathered and cleaned 1,000+ customer records, validated the resulting dataset, and prepared it for regression and cross-tabulation.

## **TECHNICAL LEADERSHIP**

---

- Led data product engineering for Gozem's governed self-service Data Hub, using application-ready datasets to power operational services, automated workflows, and decision tools beyond BI.
- Progressed from Gozem's first data hire to technical leader of a 14+ person data organization while continuing to design pipelines, review code, and own production incidents.
- Co-founded iSheero and, as Vice-Chair, designed and delivered a hands-on Google Cloud streaming module for data engineering bootcamp scholars using Pub/Sub, Dataflow, BigQuery, and Looker Studio.

## **EDUCATION & CREDENTIALS**

---

- Google Cloud Professional Data Engineer Certification, Nov 2025
- Databricks Fundamentals Accreditation, Sep 2026
- AWS Data Engineering Foundations Certificate, Aug 2026
- Apache Airflow Hands-on Certificate, Feb 2024
- MSc in Applied Statistics (Distinction), University of Abomey-Calavi, 2018
- BSc in Statistics and Econometrics, University of Abomey-Calavi, 2013

**Languages:** French (native); English (full professional proficiency)