Malcolm Simon Buluku

📍 Nairobi, Kenya | 📧 malcolmbuluku@gmail.com | 📞 +254710152496
🔗 [LinkedIn](#) | [GitHub](#)

Professional Summary.

Results-driven **Senior Data & Analytics Engineer** with 5+ years of expertise in architecting, implementing, and scaling enterprise-grade data platforms on cloud infrastructure. Currently spearheading data engineering initiatives at **Zep-Re (PTA Reinsurance)**, delivering mission-critical analytics capabilities across underwriting, claims, and actuary functions. Demonstrated success in building robust **ETL/ELT pipelines**, implementing **real-time streaming solutions**, and orchestrating **end-to-end data workflows** that drive business value. Combines technical excellence in **GCP** and modern data technologies with strong leadership skills and cross-functional collaboration to foster data-driven decision making.

Core Competencies

- **Cloud Architecture**: GCP (GCS, BigQuery, Dataflow, Pub/Sub), AWS, Azure - designing scalable, cost-optimized solutions
- **Data Engineering**: Apache Spark, Beam, batch and streaming pipelines, change data capture, schema evolution
- **Workflow Orchestration**: Apache Airflow (Cloud Composer), DAG optimization, monitoring, alerting, CI/CD integration
- **Programming & IaC**: Python, SQL, PySpark, dbt, Terraform, Docker, performance optimization
- **Data Warehousing**: BigQuery, Snowflake, data modeling, dimensional schemas, query optimization
- **Data Governance**: Lineage tracking, metadata management, data quality frameworks, access control
- **ML Operations**: Feature engineering, ML pipeline design, model deployment infrastructure, LLM integrations
- **Technical Leadership**: Mentoring junior engineers, documentation standards, Agile methodologies

Professional Experience

Data Analytics Engineer – Zep-Re (PTA Reinsurance)

Nairobi, Kenya | Jan 2025 – Present

- **Enterprise Data Platform**: Architected and implemented a modern data platform consolidating risk, claims, and partner data into a unified warehouse, reducing reporting time by 60% and enabling advanced actuarial modeling.
- **Cloud-Native ETL**: Designed and deployed resilient ETL pipelines using GCP services to integrate 10+ internal systems, REST APIs, and partner feeds into a centralized BigQuery data warehouse.
- **Data Modeling**: Created dimensional models and data marts optimized for actuarial analysis, financial reporting, and regulatory compliance, ensuring data consistency across the organization.
- **ML Data Enablement**: Partnered with data scientists to design feature engineering pipelines and automated data preparation workflows, accelerating model development cycles by 40%.
- **Workflow Automation**: Implemented sophisticated **Cloud Composer** orchestration with comprehensive error handling, monitoring, and alerting, achieving 99.9% pipeline reliability.
- **Data Governance**: Established enterprise-wide data governance practices including automated lineage tracking, quality monitoring, and PII protection for regulatory compliance.

Senior Data Engineer – Burn Manufacturing

Nairobi, Kenya | Jan 2024 – Feb 2025

- **Data Platform Modernization**: Led the architectural design and implementation of a GCP-native data platform supporting manufacturing analytics, logistics optimization, and financial reporting.
- **Real-time Processing**: Implemented near-real-time data pipelines using **Dataflow** and **Pub/Sub** to ingest IoT sensor data, enabling production anomaly detection and predictive maintenance.
- **Pipeline Orchestration**: Built sophisticated **Airflow on Cloud Composer** workflows with comprehensive error handling, versioning, and automated testing through CI/CD pipelines.
- **Performance Optimization**: Reduced data processing latency by 50% through advanced BigQuery optimization techniques and efficient Python transformation logic.
- **Team Leadership**: Mentored junior engineers, established documentation standards, and implemented Agile methodologies, resulting in a 30% increase in team velocity.

Machine Learning Engineer – Black Mental Health Matters

Remote | Aug 2021 – Dec 2023

- **Analytics Platform**: Built end-to-end data pipelines for survey and campaign analytics, enabling data-driven decision making for mental health initiatives.
- **Automated Reporting**: Designed and implemented automated ingestion from Google Sheets, forms, and APIs into BigQuery, reducing manual reporting efforts by 75%.
- **Custom Analysis**: Delivered custom Python-based analytics solutions and implemented data quality frameworks to ensure reliable impact measurement.

Data Scientist and Consultant – Chemolex

Remote | Jan 2020 – October 2022

- **Analytics Migration**: Successfully migrated legacy spreadsheet-based reporting to a scalable GCP analytics platform, enabling real-time learning metrics tracking.
- **LMS Integration**: Developed ETL workflows to centralize data from multiple learning management systems, creating unified student progress dashboards.

Education

MSc. Information Systems
United States International University – Africa

BSc. Mathematics (Statistics) and Chemistry
University of Nairobi

Certifications

- **Google Cloud Professional Data Engineer(2023)**
- **AWS Certified Solutions Architect – Associate (In progress)**
- **ALX Software Engineering Program(2023)**
- **Moringa School – Data Analytics & Agile Methodologies(2019)**

Key Projects & Achievements.

- **Enterprise Data Lakehouse (Zep-Re)**: Designed and implemented a multi-layered data lakehouse architecture (bronze/silver/gold) on GCP, enabling scalable analytics while maintaining governance and security.
- **Streaming Analytics Platform (Burn)**: Built real-time analytics capabilities using **Pub/Sub + BigQuery** for IoT sensor data, improving manufacturing efficiency by 15% through predictive maintenance.
- **ML Feature Store**: Created a centralized feature repository with automated feature engineering pipelines, reducing model development time by 35% and ensuring consistent feature definitions.
- **Cost-Optimized Infrastructure**: Implemented intelligent partitioning, clustering, and lifecycle policies in GCP, reducing data storage and processing costs by 25% while improving query performance.
- **Data Mesh Implementation**: Led the transition toward a domain-oriented data architecture, enabling domain teams to own their data products while maintaining central governance.

Availability

✅ Open to **hybrid and remote positions.**
✅ Immediate availability for interviews.

Referees:

✅ Available on Request.