LEBOHANG JOSEPH MAKHETHA

Data Engineer | 3x AWS Certified

+27 60 886 3922 | leboindata@gmail.com | Cape Town, South Africa

linkedin.com/in/lebohang-makhetha | leboindata.com | github.com/Lebo-In-Data

SUMMARY

Data engineer with 4+ years at AWS supporting enterprise clients across industries on production streaming and batch data pipelines, from root-cause investigation to architecture. Now building an open-source portfolio of batch, streaming and data quality projects while completing a BSc in Mathematics & Statistics.

SKILLS

Streaming: Kinesis Data Streams, Data Firehose, MSK (Kafka), Managed Service for Apache Flink

Storage & analytics: S3, Redshift, Glue, Athena, OpenSearch, PostgreSQL

Pipelines & quality: Airflow, Great Expectations, Grafana, Lambda, IAM, CloudFormation, Git

Modelling: Star and snowflake schemas, data marts, lineage, data quality frameworks

Languages: Python, SQL, Java, C#

Learning: dbt, Apache Spark, Iceberg, Debezium, Docker, Statistics

EXPERIENCE

Cloud Support Engineer I, Data Flow Analytics | AWS | Oct 2024 – Apr 2026 | Cape Town

- Advised enterprise clients across industries on production AWS data pipelines.
- Led root-cause investigations of high-severity pipeline failures.
- Designed multi-AZ MSK architectures handling millions of financial events per second.
- Tuned Flink parallelism, checkpointing and state backends for sustained throughput.
- Built BCBS 239-aligned data quality and lineage frameworks for financial services clients.

Cloud Support Associate, Data Flow Analytics | AWS | Feb 2022 – Oct 2024 | Cape Town

- Supported enterprise clients on MSK, Kinesis, Glue, Athena, OpenSearch, Redshift and RDS.
- Resolved latency, consumer lag and schema issues through root-cause analysis; promoted after 2.5 years.

CERTIFICATIONS

AWS certified: Solutions Architect – Associate · AI Practitioner · Cloud Practitioner

In progress: AWS Data Engineer – Associate

Other: Intro to Service Management with ITIL® 4 (LinkedIn Learning)

CLIENT PROJECTS (AWS)

Write-ups only: source code is client-confidential. Details at leboindata.com.

Multi-AZ Kafka Streaming Platform · Major African bank
Traced a throughput degradation on a payments pipeline to partition imbalance; advised on repartitioning and Flink checkpoint and state tuning. **Outcome:** throughput restored, zero data loss, multi-AZ failover validated.

Cloud Data Warehouse Design · Enterprise fintech
Guided a Redshift warehouse: star-schema model, Glue ETL review, distribution keys, sort keys and WLM queues. **Outcome:** query targets met, pipeline stable in production.

Data Quality & BCBS 239 Governance Framework · Financial services
Designed source rules, reconciliation checks, lineage standards and an OpenSearch compliance dashboard with CloudWatch alerting. **Outcome:** audit-ready, fewer recurring incidents.

Streaming Pipeline Observability Platform · Enterprise banking
Built log ingestion and an OpenSearch dashboard for throughput, consumer lag, schema errors and dead-letter queues. **Outcome:** proactive instead of reactive monitoring.

PERSONAL PROJECTS

Open source, built locally on free tools: github.com/Lebo-In-Data

Batch ETL Pipeline · **Complete**
Airflow · PostgreSQL · Python — retries, backfills and idempotent loads.

Data Quality & Observability · **In progress**
Great Expectations · Prometheus · Grafana — quality checks on the batch DAGs.

Real-Time CDC Pipeline · **Planned**
Debezium · Redpanda · DuckDB — database changes streamed with schema checks.

EDUCATION

BSc Mathematics & Statistics, University of South Africa (in progress, completing 2029)
Bachelor of Information Technology, Nelson Mandela University, 2019 – 2021
Average 79%, 19 of 27 modules with distinction.

AWARDS

Golden Key International Honour Society (top 15%, 2020 – present) · NMU Student Merit Award 2020 & 2021 · Matric Top 100 Achiever, Free State Province