THABO KUNENE

SENIOR DATA ENGINEER

Johannesburg, South Africa | +27 71 111 1630 | engineer@thabokunene.co.za | https://www.thabokunene.co.za

LinkedIn: linkedin.com/in/mntimande | **GitHub:** github.com/Klinsh | **WhatsApp:** +27 78 118 8828

PROFESSIONAL SUMMARY

Senior Data Engineer with 10+ years building Azure/AWS-native data platforms for mission-critical environments. At Teraco Data Environments, engineer high-throughput pipelines (3TB+/day) using Spark/PySpark, Kafka, ADLS Gen2, Synapse and dual orchestration (Data Factory + Airflow), delivering ~30% cost reductions and faster operational insights. Hands-on with Azure Purview for cataloguing, lineage and compliance-aligned governance. Experienced in dbt/Airflow ELT, data quality practices, and Agile/DataOps delivery. Azure DP-203 certified. Seeking to apply platform engineering and governance expertise to financial services data challenges.

SELECTED ACHIEVEMENTS

- Built and operated real-time platforms processing 3TB+/day with sub-second alerting
- Delivered ~30–40% cost reductions through platform optimisation and cloud migration
- Modernised 50+ production pipelines, improving performance by ~40–50%
- Implemented governance and lineage with Azure Purview, reducing data discrepancies by ~30%

TECHNICAL SKILLS

Azure (Core)	ADLS Gen2, Synapse Analytics, Data Factory, Azure Purview, Azure Monitor/Log Analytics
Data Engineering	Spark/PySpark, Kafka, Flink, dbt, Airflow, Informatica
Warehousing/Lake	Synapse, Snowflake, Redshift, BigQuery; dimensional modelling (star schema, Kimball)
Programming	Python (advanced), SQL (advanced), Scala/Java
Datastores	PostgreSQL, MongoDB, Cassandra, DynamoDB, Cosmos DB, Elasticsearch
Data Quality	dbt tests, custom validation frameworks, data contracts (fundamentals)
Analytics & BI	Power BI (reporting/consumption layer), Tableau, Looker
MLOps/ML	SageMaker, MLflow, XGBoost, SHAP explainability (project-based)
DevOps & Security	Git, CI/CD (Jenkins/GitHub Actions), Docker/Kubernetes, RBAC, managed identities, Azure Key Vault
Governance	Azure Purview (cataloguing, lineage, classification), POPIA, data access policies
Ways of Working	Agile (Scrum/Kanban), DataOps, Data Mesh fundamentals
Projects/POCs	Azure Databricks, Delta Lake lakehouse patterns (bronze/silver/gold)

PROFESSIONAL EXPERIENCE

Senior Data Engineer

Teraco Data Environments · Johannesburg ***Nov 2021 – Present***

Africa's largest carrier-neutral data centre operator housing critical digital infrastructure for financial services and enterprise clients.

- Architected and supported a real-time telemetry platform ingesting data from thousands of sensors (3TB+/day), enabling sub-second alerting for power, cooling and environment anomalies using Kafka + Spark/PySpark, with orchestration across Azure Data Factory and Airflow.
- Led platform modernisation using ADLS Gen2 + Synapse + dbt, improving pipeline performance by ~50% and reducing operational costs by ~30% (R10M+ savings) through compute/storage optimisation and workflow redesign.
- Implemented Azure Purview for data cataloguing and lineage, including POPIA-aligned data classification and access policies, accelerating troubleshooting and improving audit readiness; reduced data discrepancies by ~30% via validation and lineage-driven root cause analysis.
- Configured RBAC, managed identities and Azure Key Vault for ADLS Gen2 and Synapse workspaces, supporting secure data access patterns and meeting internal security and compliance requirements.
- Built forecasting and operational analytics for capacity planning and infrastructure reporting, improving prediction accuracy by ~20–25% and enabling more consistent executive reporting.
- Partnered with operations, commercial and facilities teams to define data requirements, translating business questions into pipeline designs, reporting outputs and alerting thresholds.
- Mentored 4–6 engineers and embedded DataOps practices (PR discipline, coding standards, CI/CD hygiene, dbt test coverage), increasing delivery throughput by ~25%+.

Senior Data Engineer

Jonsson Workwear · Johannesburg ***Jan 2019 – Oct 2021***

Leading South African workwear and retail company leveraging data for customer-centric operations.

- Designed and deployed a cloud-native analytics platform on AWS (Kinesis, Redshift, Snowflake, Spark), establishing a scalable foundation for retail analytics and reporting.
- Optimised storage, compute and ETL workflows, reducing costs by 25%, improving retrieval speed by 40%, and cutting processing time by 50% (~500 hours saved annually).
- Migrated 50+ pipelines to a modern stack (Kafka + Airflow + Spark) processing 2TB+/day, improving data accessibility by ~60% across the organisation.
- Built real-time streaming analytics (Flink, 100K+ events/sec), reducing latency by ~45% for inventory and customer behaviour insights.
- Collaborated with marketing, supply chain and executive stakeholders to translate business requirements into analytical pipelines and Power BI reporting dashboards.
- Employee of the Year (2020) for optimisation initiatives delivering R8.5M savings; mentored 5 engineers, improving delivery speed by ~30%.

Data Engineer / Business Analyst

Data Solutions Inc. · Cape Town ***Jun 2015 – Dec 2018***

Data consultancy specialising in analytics and digital transformation for small-to-medium enterprises.

- Automated ETL and reporting workflows (Informatica/Spark), reducing manual effort by ~70% (~300 hours/year) and data-entry errors by ~20%.
- Delivered analytics for expansion and operational improvement initiatives (R19M+), improving efficiency by ~10–15% and supporting ~15% revenue growth in key accounts.
- Worked directly with client stakeholders to gather requirements, present insights and recommend data-driven improvements.
- Received Innovation Award for advanced modelling and improved reporting design.

KEY PROJECTS

Azure Databricks Lakehouse - Personal / POC

- Built a lakehouse POC using Azure Databricks + Delta Lake with streaming ingestion patterns and curated layers (bronze/silver/gold), demonstrating streaming and batch analytics patterns with governed data layers.

- Focus areas: performance tuning, dbt tests and data quality checks, partitioning strategy and reproducible deployment notes. GitHub: github.com/Klinsh/lakehouse-poc

Real-Time Streaming Platform - Teraco / Jonsson

- Kafka/Flink streaming system enabling near real-time analytics; ~45% latency reduction. Stack: Kafka, Flink, Spark/PySpark, Elasticsearch.

Predictive Churn Engine - Personal / Jonsson

- Built an XGBoost churn model with retraining automation and SHAP explainability; achieved ~18–22% churn reduction on test/sample data. Stack: Python, SageMaker, S3, Lambda, dbt.

EDUCATION & CERTIFICATIONS

B.Sc. Computer Science - Andrews University (2011)

Certifications

- Microsoft Certified: Azure Data Engineer Associate (DP-203) - 2023
- Google Cloud Professional Data Engineer - 2022
- AWS Certified Data Analytics – Specialty - 2021
- Deep Learning Specialization (deeplearning.ai) · PySpark (LinkedIn Learning) · ML Explainability (Kaggle)
- *In progress: Azure AI Engineer (AI-102) · AWS ML Specialty*
- *Databricks Data Engineer Associate - exam scheduled Q3 2025*

PUBLICATIONS, OPEN SOURCE & COMMUNITY

Publications / Talks

- “Scalable Fraud Pipelines with Delta Lake” - Data + AI Summit Africa (2024)
- “Deep Learning for Real-Time Predictive Maintenance” - Keynote, International Conference on Data Science and Engineering (2021)
- “Building Resilient Data Pipelines for High-Velocity IoT Data” - Big Data Analytics Symposium, Johannesburg (2022)
- “Optimizing ETL with Cloud-Native Technologies” - DataOps Africa Conference (2023)

Open Source & Community

- Apache Airflow PRs (pipeline stability); dbt adapters (10+ companies, 500+ downloads)
- Johannesburg Data Engineering Meetups (organiser/speaker); PyCon ZA speaker

LANGUAGES

English (Professional) **Zulu** (Native) **French** (Intermediate) **Setswana** (Basic)

South African citizen · Notice period: 1 month · Available for background screening

Interests: Fintech/RegTech, sustainable data infrastructure, engineering mentorship

Portfolio: https://www.thabokunene.co.za/Portfolio