

Peter Opapa

✉ Email: opapapeter82@gmail.com | 📞 Phone: +254 743695612 | 💼 LinkedIn: linkedin.com/in/peter-opapa

🌐 Portfolio Website: www.peteropapa.tech | 💻 Github: github.com/Peter-Opapa

About

With over one year of hands-on project experience, I specialize in building efficient data engineering solutions using Big Data tools and cloud technologies. My expertise in Python, SQL, Azure, Spark, and Databricks enables me to design scalable data pipelines that enhance data acquisition, transformation, and storage performance.

Experience

Electronics Engineering Intern (IoT), Ubuntu Water Hub

May 2025 – August 2025

- Assisted in the design and simulation of IoT devices used for real-time water pump monitoring and control.
- Evaluated prototype performance, focusing on sensor data accuracy, reliability, and signal integrity.
- Analyzed real-time data generated by devices (flow rate, status, pressure) to inform system design and evaluation.

Projects

This is a selection of my key projects. For a full list of all the projects and their details, please visit my [portfolio](#).

Azure Data Engineering Project: Enterprise-Scale Medallion Architecture

- Engineered a production-ready Medallion Architecture (Bronze, Silver, Gold) using **Azure Data Factory**, **Databricks (PySpark)**, and **Azure Data Lake Storage**.
- Loaded transformed data into **Azure Synapse Analytics** for final analysis and reporting in **Power BI**.

MovieLens Data Pipeline - ADLS Gen2 + dbt + Snowflake

- Constructed a modern ELT pipeline extracting data from **Azure Data Lake Storage Gen2**.
- Loaded raw data into **Snowflake** and used **dbt Core** for complex data modelling and transformations.
- Created a dimensional model optimized for analytics, complete with auto-generated documentation.

Weather Data Streaming Pipeline

- Ingested real-time weather data from external APIs using Python scripts hosted in **Azure Functions** and **Azure Databricks**.
- Published the raw data stream to **Azure Event Hubs** for scalable, real-time ingestion.
- Processed and analyzed the live data stream using **Azure Stream Analytics** to identify key metrics.
- Visualized real-time insights and dashboards using **Power BI**.

Real-time Data Pipeline (Kafka & Spark)

- Built a real-time streaming pipeline using **Apache Airflow** for orchestration and **Kafka** for data decoupling.
- Utilized **Apache Spark** for stream processing and stored enriched data in **Cassandra** and **PostgreSQL**.
- Containerized the entire system using **Docker** for portability and scalability.

Financial Transactions Pipeline (Flink & Kafka)

- Developed a real-time analytics pipeline using **Kafka** for ingestion and **Flink SQL** for event-time aggregations.
- Stored the final results in **PostgreSQL** and monitored system health in real-time with **Datadog**.

Education

University of Nairobi, Bsc. Electrical and Electronics Engineering Sep 2021 - Sep 2026

Certifications

- **Confluent Certified Data Streaming Engineer** - Confluent
- **MongoDB Certified Associate Developer** - MongoDB
- **Oracle Cloud Infrastructure Certified Architect** - Oracle
- **Oracle Cloud Infrastructure Certified Generative AI Professional** - Oracle
- **Snowflake-Data Engineering** - Snowflake
- **Snowflake-Data Warehouse** - Snowflake
- **Azure Databricks Platform Architect** - Databricks
- **Databricks Fundamentals** - Databricks
- **dbt Fundamentals** – dbt Labs

Skills

- **Cloud Data Platforms:** Azure (ADF, ADLS, Synapse), Databricks, Snowflake, Oracle Cloud (OCI)
- **Big Data Frameworks:** Apache Spark, Apache Flink, Apache Kafka, Apache Airflow, dbt Core
- **Programming & Databases:** Python, SQL, PostgreSQL, MongoDB, MySQL
- **DevOps & Tools:** Docker, Git & GitHub, GitHub Actions, Terraform
- **Analytics & Visualization:** Power BI, Excel

Referee

Available on Request.