

Samuel Kapesa



in/samuel-kapesa-jr-5b3293201

SUMMARY

- Data analyst transitioning into data engineering with a strong foundation in Python, SQL, and cloud technologies, specializing in building scalable data pipelines, automating workflows, and managing large datasets. Experienced in cloud platforms like Google Cloud (GCP) and AWS. Proven ability to optimize data operations, ensure data integrity, and deliver real-time analytics to drive decision-making. Adept at both batch and stream data processing with a passion for developing efficient, high-performance solutions.

EXPERIENCE

Data Analytics Consultant

The Center For Infectious Disease Research - CIDRZ

August 2023 - Present, Lusaka, Zambia

- Automated **data pipelines**, reducing manual data processing time by 30%, which improved the data flow between systems and minimized errors.
- Led a project that integrated **Apache Kafka** for real-time data ingestion and stream processing, which enhanced data availability for analysis.
- Built ETL processes to clean, transform, and load data into **Google BigQuery**, ensuring efficient data storage and retrieval for research teams.
- Created real-time **Power BI dashboards** and reporting tools that delivered actionable insights based on processed data, boosting operational efficiency by 25%.

Data Analysis Intern

CIDRZ

February 2023 - July 2023, Lusaka, Zambia

- Developed scalable data workflows using **Python** and **SQL** to handle large epidemiological datasets, ensuring data consistency and reliability.
- Merged multiple datasets into centralized repositories, applying **data governance principles** to maintain high data integrity across the organization.
- Deployed machine learning models in **GCP** for predicting disease spread, utilizing **BigQuery** for querying historical data and forecasting trends.
- Built a real-time **data ingestion** system using **Python** and cloud storage solutions, enabling rapid access to data for health analytics teams.

Social Media Marketing Intern

Asia Pacific University

June 2022 - December 2022, Lusaka, Zambia

- Utilized data-driven strategies to optimize social media campaigns, applying **SQL** to track and analyze engagement metrics for performance improvement.
- Implemented targeted advertising with a 25% reduction in cost per click, demonstrating the ability to leverage data for optimizing business decisions.

Personal Assistant & Data Entry

Premier Medical Corporation

January 2017 - September 2018, Lusaka, Zambia

- Managed large datasets, performing data entry with 98% accuracy, which maintained data integrity and improved workflow efficiency by 30%.
- Organized and structured data for operational use, laying a foundation for future automation and data processing improvements.

EDUCATION

Bachelors Of Science Computer Science

Minor in Data Analytics • Asia Pacific University • Kuala Lumpur, Malaysia • 2022 • 3.0

SKILLS

- **Python (pandas, numpy, PySpark)**: Skilled in building scalable ETL processes for large datasets.
- **SQL (Postgres, MySQL, BigQuery)**: Expertise in querying, optimizing, and managing complex datasets for high-performance data processing.
- **Google Cloud Platform (GCP)**, **AWS** and **OCI** for deploying scalable machine learning models and data pipelines.
- Proficient in **Apache Kafka** for stream processing and real-time data integration.

- Developed end-to-end **data pipelines** for cleaning, transforming, and loading large datasets, improving efficiency and reducing processing time.
- Automated data workflows using **Python** and **SQL**, reducing manual work by up to 30% and enabling faster decision-making.
- Experience with designing and maintaining **data warehouses** in **PostgreSQL** and **SQL Server**, ensuring performance and reliability for analytics workloads.
- Extracted large volumes of data from web sources using **Scrappy**, **Beautiful Soup**, and **Selenium** to feed into data pipelines for further analysis.
- Created **Power BI dashboards** for real-time analytics, providing stakeholders with actionable insights.
- Automated reporting systems using **Python**, delivering key metrics and financial reports in real time.