

JOHN GITAU NJENGA

[Johngnjenga99@gmail.com](mailto:Johngnjenga99@gmail.com) +254799238443 - Nairobi, Kenya

## Profile

I am a skilled and passionate data engineer with a strong background in improving and building robust data systems, storage, processing platforms, visualization, and data analysis. I take immense pleasure in designing and constructing large-scale batch and real-time data systems.

## Experience

DATA ENGINEER

University of California San Francisco (UCSF) Nairobi – Global health security.

May 2023 – Present

## Responsibilities and Accomplishments

- Collaborated closely with stakeholders, funders, and government officials to gather and refine data requirements.
- Designed, implemented, and maintained high-performance surveillance databases using PostgreSQL, focusing on data accuracy and integrity.
- Created complex Extraction, Transformation, and Loading (ETL) queries using SQL stored procedures and Apache Spark 3.3.0 to extract and process data.
- Utilized Superset to create intuitive and interactive dashboards for visualizing public health data, enabling stakeholders to derive actionable insights.
- Ensured data quality by implementing rigorous data validation processes and conducting regular data quality audits.

## Achievements

- Developed a health insurance billing dashboard based on Apache Superset and PostgreSQL which was integrated with the Rwanda EMR under OpenMRS.
- Developed SQL ETL scripts and Power Bi dashboards for OneHealth which focuses on detecting the next pandemic before it really happens by collecting and analyzing animal, human and environment data.

- Automated data quality assurance processes that reduced human error and improved data accuracy by 15% using Great Expectation tool. This enhancement ensured that data used for critical decision-making was consistently reliable and error-free.
- Facilitated Apache Superset hackathon held in Mombasa to sensitize Zambia MOH and partners on superset best practices and how to produce informative dashboards.

## DATA ANALYST

Global Programs For Research And Training Affiliate Of The University Of California San Francisco (UCSF) - Health informatics - Nairobi

April 2021 – May 2023

### **Responsibilities and Accomplishments**

- Utilized SQL to write ETL programs, extracting data from source databases and transforming it into a dimensional model, including base, derived, and final dimensions and facts.
- Developed database views and other objects to facilitate querying, reporting, and analysis.
- Produced comprehensive database documentation, including procedures and data dictionary definitions.
- Designed and developed Apache Superset and Power BI dashboards for the Ministry of Health Kenya, enabling real-time tracking of HIV spread and patient behavior analysis.
- Implemented Apache spark and Apache Airflow to create and manage complex data pipelines, automating ETL jobs and enhancing data pipeline efficiency.
- Led the Kenyan team in utilizing Palantir Foundry, organizing technical meetings between organizations.
- Collaborated with epidemiologists to provide data, tools, and support for data-driven research and analysis.
- Established and maintained data quality checks and validation procedures to ensure data integrity and accuracy in data warehouses.
- Developed machine learning models to predict the health status of HIV patients in Kenya, aiding in decision-making.

### **Key Achievement**

- Provided Power Bi dashboards from intricate surveillance databases to the Jordan Ministry of Health (JMOH) for their signal/event based data, enabling more effective public health decisions.
- Provided technical assistance to partners (Rwanda MOH, Jembi) on integrating SQL ETLs and Apache Superset dashboards into OHRI (OpenMRS).
- Managed the HIV viral load database, ETLs and dashboards, which played a vital role in monitoring the effectiveness of antiretroviral therapy (ART) and reducing the spread of the virus, ultimately saving lives.

## Education

Bachelor of Science In Information Technology

KCA University, Nairobi, Kenya

Sep 2017 - April 2021

Second class honors - upper division

## Skills

### *Database Management Systems*

- PostgreSQL
- Microsoft SQL Server
- MySQL
- MongoDB

### *Big Data Technologies*

- Apache Spark
- Hadoop
- Apache airflow

### *Business Intelligence*

- Apache Superset
- Power BI

### *Data Warehousing and Data Modeling*

- Expertise in data modeling and schema design - transforming data into dimensional models consisting of both dimensions and facts using the Kimball's approach.
- ETL development

### *Programming Languages*

- Python