

STI07-ETI-021

Introducing a Data Pipeline for The Tahmo Network

Austin Kaburi^a, Jason Kabi^a and Ciira Maina^a

^aCentre for Data Science and Artificial Intelligence (DSAIL), Dedan Kimathi University
of Technology, Private Bag-10143, Nyeri, Kenya

Email: kibaara.austin20@students.dkut.ac.ke

Abstract

The Trans-Africa Hydro-Meteorological Observatory (TAHMO) is dedicated to alleviating the data scarcity that has long hampered African farmers' decision-making processes. With the ambitious objective of establishing a network of 20,000 weather stations across Africa, TAHMO currently operates 700 weather stations across 25 African countries. To manage this ever-expanding network efficiently, this poster introduces a data pipeline built on Google Cloud. The data pipeline leverages serverless architectures, Cloud Functions, and App Engine to reduce operational costs. Its primary goal is to collect, store, and analyze data from these weather stations, with a particular focus on precipitation data. The process involves data extraction, precipitation and the TAHMO's flags from the regression model, and integration of ground truth data from on-site technicians. A cloud scheduler triggers the data extraction and loading process on Google Cloud Storage. Dataflow processes this information in batches to ensure conformity with the warehouse's schema. The result is a continuous reporting system

that enables real-time data analysis. This data pipeline simplifies data access, eliminating the need for manual data extraction and transformation. Future work will involve integrating different models to enhance the quality of data provided to farmers, thereby improving agricultural decision-making in Africa.

Keywords: TAHMO, Serverless Architecture, Data extraction, Agricultural decision-making, Data Pipeline.