

ABSTRACT

The Faculty of Informatics produces various types of important data, such as research data and community service data. However, its management still faces challenges, including separated data storage and difficulties in analyzing both structured data and semi-structured data. This research discusses about the design and implementation of frameworks based on data lakehouse architecture using Google Cloud Platform (GCP) as a solution for more efficient data integration and data processing. Data lakehouse is chosen based on its combination of data lake and data warehouse advantages, which enables data analysis according to the domain problem. Two frameworks were designed: one for structured data utilizing Vertex AI, BigQuery, and a chatbot, and another one for semi-structured data using a chatbot powered by Gemini 2.0 Flash. The results shows that the frameworks are flexible enough to handle contextual analysis, so that it supports data-driven decision-making.

Keywords: *data lakehouse, data lake, data warehouse, framework, data analysis, Google Cloud Platform*