

ABSTRAK

Universitas Telkom menghadapi tantangan dalam mengelola dan mengintegrasikan data dari berbagai sumber dengan format yang beragam. Penelitian ini bertujuan untuk merancang dan mengimplementasikan *framework data ingestion* yang dapat menangani data yang beragam ke dalam sistem *data lakehouse* perguruan tinggi berbasis Google Cloud Platform (GCP). *Framework* ini memanfaatkan layanan seperti Cloud Storage, BigQuery, Pub/Sub, Eventarc, dan Cloud Function untuk mengumpulkan, mengolah, serta mengintegrasikan data secara otomatis. Dengan pendekatan *pipeline*, *framework* ini dirancang sebagai jalur penghubung data berbasis metrik indikator kinerja akreditasi pada program studi S1 Informatika, mencakup kinerja mahasiswa, pendidikan, penelitian, pengabdian kepada masyarakat, serta luaran dan capaian tridarma. Sebagai bagian dari proses integrasi, *framework* ini juga membangun *data warehouse* untuk menggabungkan seluruh data akreditasi untuk analisis berbagai metrik.

Hasil dari penelitian ini adalah *framework data ingestion* yang dapat mengintegrasikan data dari berbagai sumber berdasarkan metrik akreditasi ke dalam *data lakehouse* secara otomatis, sehingga meningkatkan efisiensi dalam pengelolaan data akreditasi dan mendukung analisis kinerja indikator akreditasi dalam pengambilan keputusan berbasis data. Hasil evaluasi menunjukkan bahwa peningkatan alokasi CPU mempengaruhi kecepatan *ingestion*, di mana peningkatan dari 333 millicore ke 1 vCPU membuat proses *ingestion* menjadi jauh lebih cepat. Selain itu, jumlah data berhubungan langsung dengan penggunaan memori, di mana batasan 512 MiB menyebabkan masalah *Out of Memory* (OOM) untuk memproses sekitar 150.000 *record*, sehingga diperlukan konfigurasi memori yang lebih besar.

Kata Kunci: Keberagaman sumber dan format data, Data Ingestion, Data Integration, Data Lakehouse, Google Cloud Platform.