

# Extensible High-Throughput Streaming Platform Berbasis Kafka untuk Pemrosesan Data dari Media Sosial = Kafka-Based Extensible High Throughput Streaming Platform for Data Processing from Social Media

Aldi Naufal Fitrah, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=9999920531689&lokasi=lokal>

---

## Abstrak

Dengan bertambahnya jumlah pengguna internet seiring waktu, bertambah pula jumlah data yang dihasilkan dari aktivitas yang berasal dari media sosial. Data tersebut dapat digunakan untuk berbagai keperluan, salah satunya untuk keperluan moderasi supaya aktivitas pengguna internet tetap tunduk pada hukum yang berlaku. Namun, diperlukan cara yang optimal untuk melakukan proses tersebut mengingat data yang jumlahnya sangat besar. Penelitian ini mengajukan sebuah platform yang dapat menjadi salah satu pilihan untuk memproses data media sosial yang berjumlah besar tersebut. Rancangan platform pada penelitian ini ditujukan untuk dapat memiliki throughput yang besar. Selain itu, platform dirancang untuk dapat dimodifikasi demi memenuhi berbagai kebutuhan. Karenanya, aspek extensibility juga menjadi perhatian utama dalam proses pengembangan platform. Kedua tujuan utama dalam pengembangan platform ini dapat diwujudkan dengan bantuan sebuah kluster Apache Kafka yang membuat platform memiliki sifat loosely-coupled dan juga extensible. Dengan berpusat pada kluster Apache Kafka, proses pengolahan data yang ada dapat dilakukan secara paralel, dan terbukti dapat meningkatkan throughput dari platform secara keseluruhan. Sebagai pembanding, penelitian ini diuji coba dengan suatu skenario bersama dengan platform Tweetream yang dikembangkan oleh Susanto (2022). Hasil dari uji coba tersebut membuktikan bahwa platform pada penelitian ini dapat mengungguli Tweetream.

.....As the number of internet users increases over time, so does the amount of data generated from activities originating from social media. This data can be used for various purposes, one of which is for moderation purposes so that the activities of internet users remain subject to applicable laws. However, an optimal way to do the process is needed considering the huge amount of data. This research proposes a platform that can be one of the options for processing large amounts of social media data. The design of the platform in this research is intended to have a large throughput. In addition, the platform is designed to be modifiable to meet various needs. Therefore, extensibility is also a major concern in the platform development process. These two main objectives in the development of the platform can be realized with the help of an Apache Kafka cluster that makes the platform loosely-coupled and extensible. By centering on the Apache Kafka cluster, the data processing can be done in parallel, which has been proven to increase the throughput of the platform as a whole. For comparison, this study was tested in a scenario with the Tweetream platform developed by Susanto (2022). The results of the test proved that the platform in this study can outperform Tweetream.