

## ABSTRACT

Twitter is one of the social media that contributes data with a considerable number of their tweets. A large amount of data and continue to grow can make the amount of data very large or can be referred to as big data. This large amount of data can be used to influence a decision and to speed up the discovery. Storage and processing of these data can be done with cloud computing because cloud computing has internet network facilities in a broad scope. For example, like the Google Big Query dan Elasticsearch frameworks that can process the big data. Benchmarking is a process used to measure and compare performance against activities. Benchmarking performance on big data was developed to evaluate and compare the performance of a system and the architecture of the big data.

This study aims to obtain the level of time effectiveness between processing speed and analyzing a data between Google Big Query and Elasticsearch. The attributes used in the comparison are Query Time and Total Time for platform performance speed. Other attributes are Request Sent, Waiting (TTFB) and Content Download for the request and response times provided by the browser. Using 5,000 tweets retrieved from Twitter, the Google Big Query and Elasticsearch platforms were compared using three models. The first model uses a SELECT query, and the second model uses a query with the conditions user\_location and user\_bio IS NOT NULL and user\_follower\_count and user\_created\_count  $\geq 1$  and sorted by the lowest user\_created\_at. While in the third model, the tweet\_type condition is a reply and the description has #COVID19. The results show that Google Big Query has a stable and efficient rate based on the speed of query processing time performance than the Elasticsearch.

**Kata kunci** – *Benchmarking, Google Big Query, Elasticsearch, Twitter, Tweet*

## INTISARI

Twitter merupakan salah satu media sosial yang menyumbangkan data dengan jumlah yang cukup banyak dari cuitannya. Dengan banyaknya jumlah data tersebut dan akan terus menerus bertambah dapat menjadikan jumlah data menjadi yang sangat besar atau bisa disebut sebagai *big data*. Jumlah data besar tersebut dapat digunakan untuk mempengaruhi suatu keputusan dan untuk mempercepat suatu penemuan. Penyimpanan dan pemrosesan data tersebut dapat dilakukan dengan *cloud computing*, karena *cloud computing* memiliki fasilitas jaringan internet dalam cakupan yang cukup luas. Seperti pada *platform Google Big Query* dan *Elasticsearch* yang mampu memproses *big data* tersebut. *Benchmarking* merupakan salah satu proses yang digunakan untuk mengukur dan membandingkan suatu kinerja terhadap aktivitas. *Benchmarking* yang dilakukan pada *big data* dikembangkan untuk mengevaluasi dan membandingkan kinerja dari suatu sistem dan arsitektur dari *big data* tersebut.

Penelitian ini bertujuan untuk mendapatkan tingkat efektifitas waktu antara kecepatan dalam pengolahan dan menganalisa suatu data antara *Google Big Query* dan *Elasticsearch*. Atribut yang digunakan dalam perbandingan adalah *Query Time* dan *Total Time* untuk kecepatan kinerja *platform*. Atribut lainnya adalah *Request Sent*, *Waiting (TTFB)* dan *Content Download* untuk waktu *request* dan *response* yang diberikan *browser*. Menggunakan 5.000 cuitan yang didapatkan dari *Twitter*, *platform Google Big Query* dan *Elasticsearch* dibandingkan menggunakan tiga model. Yang pertama menggunakan kueri *SELECT*, dan model kedua dengan menggunakan kueri dengan kondisi *user\_location* dan *user\_bio IS NOT NULL* serta *user\_follower\_count* dan *user\_created\_count*  $\geq 1$  dan diurutkan berdasarkan *user\_created\_at* yang terendah. Sedangkan pada model ketiga ditambahkan kondisi *tweet\_type* adalah *reply* dan *description* memiliki #COVID19. Hasil penelitian ini menunjukkan *Google Big Query* memiliki tingkat yang stabil serta efisien berdasarkan kecepatan kinerja waktu pemrosesan kueri dibandingkan *Elasticsearch*.

**Kata kunci** – *Benchmarking, Google Big Query, Elasticsearch, Twitter, Tweet*