

Intisari

Analisis media sosial Twitter dapat menunjukkan *rating* seseorang, layanan, atau suatu produk di mata pengguna Twitter. Sebagai salah satu media sosial dengan pengguna terbanyak di dunia, Twitter menyediakan API yang memudahkan pengambilan data Twitter secara *real time*. Analisis ini menjadi penting karena dapat dimanfaatkan sebagai dasar penarikan keputusan. Elasticsearch adalah *tools* yang mampu melakukan analisis *big data*; termasuk data Twitter yang *realtime*. Dari sekian banyak *tools input* yang dimilikinya, Elasticsearch memiliki dua *tools* yang dapat menginputkan data Twitter dari Twitter API ke Elasticsearch. Cara pertama adalah melalui Twitter River dan cara kedua adalah melalui Logstash. Faktor *input* adalah faktor yang sangat mempengaruhi *output*. Keakuratan data *input* serta efisiensi pengolahan serta penyimpanan data sangat penting dalam menunjang sebuah sistem yang menangani *big data*. Dalam rangka mengoptimalkan kinerja sistem analisis media sosial Twitter dengan Elasticsearch dan Kibana, diperlukan sebuah penelitian yang membandingkan kedua *tools input* tersebut.

Penelitian dilakukan dengan memonitor dua buah server HPC yang melakukan *crawling* data dari Twitter API secara bersamaan. Server HPC pertama adalah server Elasticsearch yang menggunakan *input* Logstash, sedangkan server HPC kedua menggunakan *input* Twitter River. Parameter pembanding yang digunakan adalah beban pemrosesan CPU, penggunaan RAM, penggunaan *disk*, jumlah data Twitter (*twit*) yang diinputkan, serta jumlah *field* yang diinputkan pada Elasticsearch.

Hasil penelitian menunjukkan bahwa Logstash menggunakan lebih banyak sumber daya dibandingkan Twitter River termasuk 0,99% beban pemrosesan CPU dan 7,2% penggunaan RAM per hari, 112 MB penggunaan disk per minggu, dan 497 *field* lebih banyak per *twit*. Sebaliknya, Twitter River menginputkan 191 *twit* lebih banyak dibandingkan Logstash dalam satu minggu.

Kata kunci : Elasticsearch, Logstash, Twitter River, Analisis Media Sosial, *Big data*

Abstract

Social media analysis of Twitter can show a rating of someone, a service, or a product from Twitter user's perspective. As one of social media with highest number of user in the world, Twitter provide an API that allow us to observe and take Twitter data realtime. This analysis is important because it could be used as a fundamental reference to take decision. Elasticsearch is a tool that have the ability to do big data analysis. There are two ways to inputing Twitter data to Elasticsearch. The first way is through Twitter River and the second way is through Logstash. The input factor is an important factor that influencing output of system. The accuracy and efficiency of data inputed and the way of storing data is really important to support a system of big data. To optimizing the performance of social media analysis with Twitter data using Elasticsearch, there is a need to compare and find which is the best inputing tools in the case of inputing Twitter data to Elasticsearch.

This research is done by monitoring two HPC servers that is crawling data from Twitter API simultaneously. The first HPC server is an Elasticsearch server using Logstash input, and the second HPC server is using Twitter River input. Parameters used to compare both techniques are CPU process load, RAM used, disk used, Twitter input data, and amount of input fields.

The result of this research shows that Logstash is using more resource than Twitter River including 0.99% more CPU process load and 7.2% more RAM usage per day, 112 MB more disk usage per week, and 497 more fields per tweet. In other hand, Twitter River inputs 191 more tweets than Logstash in a week.

Keywords : *Elasticsearch, Logstash, Twitter River, Social Media Analysis, Big data*