

INTISARI

ANALISIS KINERJA INDEXING DAN SEARCHING MENGGUNAKAN PENGATURAN SHARD DAN STRUKTUR INDEKS PADA ELASTIC SEARCH

Ratri Abdatush Sholihah

14/373763/PA/16422

Proses *indexing* dokumen digunakan untuk keperluan pencarian dokumen atau publikasi ilmiah. Publikasi ilmiah berupa dokumen skripsi bagian abstrak membutuhkan proses *indexing* dalam pengelolaannya. Kinerja proses *indexing* berupa ukuran indeks dan kecepatan waktu pencarian dipengaruhi oleh banyak faktor, yaitu arsitektur sistem, struktur indeks, spesifikasi mesin dan karakteristik data.

Penelitian ini mengeksplorasi peran *shard* dan struktur indeks terhadap kinerja proses *indexing* dan *searching* pada *Elastic Search*. *Elastic Search* merupakan mesin pencari *open source* terdistribusi yang mengakomodasi kebutuhan penyimpanan data besar dan kecepatan pencarian dengan performa yang baik. Tahapan yang dilakukan adalah akuisisi data, ekstraksi data menggunakan *Apache Tika* dan pengaturan struktur indeks, pengaturan jumlah *shard* pada *single node* serta pengujian *indexing* dan *searching*. Pengaturan struktur indeks memisahkan konten dokumen menjadi *field* judul, nama penulis, nim, abstrak dan kata kunci. Pengaturan jumlah *shard* menggunakan 1 *shard*, 5 *shard*, 10 *shard* dan 15 *shard*.

Hasil pengujian menunjukkan bahwa pengaturan jumlah *shard* dan pengaturan struktur indeks berpengaruh pada kinerja *indexing* dan *searching* pada *Elastic Search* berdasarkan parameter ukuran indeks dan waktu pencarian. Ukuran indeks yang dihasilkan berbanding lurus dengan jumlah *shard* yang digunakan. Pengaturan struktur indeks membuat ukuran indeks menjadi lebih kecil. Waktu pencarian berada pada titik optimal pada penggunaan 15 *shard* pada ekstraksi data yang struktur indeksnya telah diatur.

Kata kunci: *shard*, struktur indeks, *indexing*, *searching*, *Elastic Search*

ABSTRACT

ANALYSIS ON INDEXING AND SEARCHING PERFORMANCE USING SHARD AND INDEX STRUCTURE IN ELASTIC SEARCH

Ratri Abdatush Sholihah

14/373763/PA/16422

Indexing documents used for searching documents or scientific publications. Abstract sections of Scientific publications require indexing process management. Index size and search times of indexing performance is influenced by many factors, including system architecture, the index structures, the engine specifications and the data characteristics.

This research explores the role of index structure and shards influence indexing and searching performance in Elastic Search. Elastic Search is a open source search engine that accommodates the needs of distributed big data storage and search speed with good performance. Methods being taken is the data acquisition, data extraction using Apache Tika, setting the index structure, setting the number of shard on a single node, indexing and searching. The configuration of the index structure separate content into the fields that contain of document title, author name, nim, abstract and keywords. Setting the amount of shard using 1 shard, 5 shards, 10 shards and 15 shards.

The results show that the number of shard and configuration of the index structure affected the performance indexing based on the parameter of the index size and search times. The configuration of the index structure makes the index size become smaller. The search times get optimum value when using 15 shard and data extraction that has been configured.

Keywords: shard, index structure, indexing, searching, Elastic Search