

Performance evaluation of nosql-cassandra over relational data store-mysql for bigdata

Sangeeta Gupta, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=9999920521644&lokasi=lokal>

Abstrak

The massive amounts of data collected from numerous sources like social media, e-commerce websites are a challenging aspect for analysis using the available storage technologies. Relational databases are a traditional approach of data storage more suitable for structured data formats and are constrained by Atomicity, Consistency, Isolation, and Durability (ACID) properties. In the modern world, data in the form of word documents, pdf files, audio and video formats are unstructured. Therefore, tables and schema definition are not a major concern, Relational databases, such as Mysql, may not be suitable to serve such Bigdata. An alternate approach is to use the emerging Nosql databases. In this work, a comprehensive performance and scalability evaluation of large web collection data in data stores, such as Nosql-Cassandra and relational-Mysql, is presented. These systems are evaluated with data and workloads that can be found related to Bigdata, yielding scalability of applications. The insights presented in this work serve not only for performance and scalability, but also as lessons learned and experiences relating to the configuration complexity and evaluation in sorting out the complex queries of what data storage can be used on which usage cases for large data sets. The results show how the Bigdata collected across the Web with billions of records generating continuously are poorly evaluated with Mysql in terms of 'write' operations, but how these perform well with Nosql-Cassandra. This paper yields a new approach which is unique in representing Nosql-Cassandra's poor performance in retrieval of records and disk utilisation with ever-increasing loads. The results presented in this paper show an improvement in 'read' performance with the proposed architecture and configuration over Mysql, achieving cost saving benefits to any organisation willing to use Nosql-Cassandra for managing Bigdata for heavy loads.