

Resumen

En la actualidad, las bases de datos relacionales son mucho más utilizadas frente a las no relacionales, debido al gran crecimiento de Internet, las organizaciones necesitan de un gestor de bases de datos que brinde eficiencia tanto en constancia, durabilidad, rendimiento y estabilidad frente a cualquier área empresarial, capaz de almacenar grandes cantidades de datos. Cassandra es altamente comparado con diferentes gestores de base de datos NoSQL, debido a su alta posibilidad de no fallos al almacenar una gran cantidad de datos, por otra parte, tenemos a PostgreSQL, un gestor de base de datos relacional, reconocido por brindar confiabilidad e integridad de datos gracias a la arquitectura que conlleva. El objetivo del siguiente proyecto es mediante un estudio comparativo, analizar el rendimiento de dos bases de datos PostgreSQL y Cassandra; en escritura, lectura y proceso de datos en presencia de una arquitectura en común tanto en hardware como software, realizado bajo las operaciones CRUD, las mismas que estarán basadas en un millón de datos y 16 hilos de trabajo frente al framework YCSB, herramienta que permite el análisis de rendimiento en seis escenarios con distintas características, soportado por los gestores de base de datos lineal (PostgreSQL) y no lineal (Cassandra). Los resultados demostraron a PostgreSQL con mucho más desempeño que Cassandra ante cada escenario, en vista a la fase de carga de datos y fase transaccional.

Palabras claves: PostgreSQL, Cassandra, Rendimiento, Escalabilidad, SGBD, YCSB.

Abstract

Nowadays, relational databases are much more used than non-relational ones, due to the great growth of the Internet; organizations need a database manager that provides efficiency in terms of consistency, durability, performance and stability to any business area, capable of storing large amounts of data. Cassandra is highly compared to different NoSQL database managers, due to its high possibility of no failures when storing a large amount of data, on the other hand, we have PostgreSQL, a relational database manager, recognized for providing reliability and data integrity thanks to its architecture. The objective of the following project is through a comparative study, to analyze the performance of two databases PostgreSQL and Cassandra; in writing, reading and processing data in the presence of a common architecture in both hardware and software, performed under CRUD operations, which will be based on a million data and 16 threads against the YCSB framework, a tool that allows performance analysis in six scenarios with different characteristics, supported by linear database managers (PostgreSQL) and non-linear (Cassandra). The results showed PostgreSQL with much better performance than Cassandra in each scenario, in view of the data load phase and transactional phase.

Keywords: PostgreSQL, Cassandra, Performance, Scalability, DBMS, YCSB.