

## **Comparative study for load management of HBase and Cassandra distributed databases in big data**

### **ABSTRACT**

The advancement in cloud computing, the increasing size of databases and the emergence of big data have made traditional data management system to be insufficient solution to store and manage such large-scale data. Therefore, there has been an emergence of new mechanisms for data storage that can handle large-scale data. NoSQL databases are used to store and manage large amount of data. They are intended to be open source, distributed and horizontally scalable in order to provide high performance. Scalability is one of the important features of such systems, it means that by increasing the number of nodes, more requests can be served per unit of time. Distribution and scalability are always companied with load management, which provides load balancing of work among multiple nodes. Load management efficiency varies from system to another according to the used load balancing technique. In this study, HBase and Cassandra load management with scalability will be evaluated as they are the most popular NoSQL databases modeled based on Big Table. In particular, this paper will compare and analyze the load management for the distributed performance of HBase and Cassandra using standard benchmark tool named Yahoo! Cloud Serving Benchmark (YCSB). The experiments will measure the performance of database operations with a different number of connections using different numbers of operations, database size, and processing nodes. The experimental results showed that HBase can provide better performance as the number of connections increase in the presence of horizontal scalability

**Keyword:** Big Data; BigTable; Cassandra; HBase; Load management; YCSB