Optimizing Schema and Indexing Strategies for Large Databases: A Comparative Study of SQL, NoSQL, and HTAP Systems

AU Nuwan[1#], WMKS Ilmini[2], and N Wedasinghe[3]

[1,3]Department of Information Technology, Faculty of Computing, General Sir John Kotelawala Defence University
[2]Department of Computer Science, Faculty of Computing, General Sir John Kotelawala Defence University

[#]39-bit-0033@kdu.ac.lk

Abstract

Most large-scale modern applications are increasingly in need of database management systems that are effective and scalable, meeting increased data complexity and volume continuously. This study explores the programming and indexing methodologies of SQL, NoSQL, and Hybrid Transactional/Analytical Processing (HTAP) systems, assessing their suitability for large-scale databases. The objective is to compare the structured, transactional strengths of SQL with the scalability and flexibility of NoSQL for unstructured data, and to evaluate HTAP systems for handling both transactional and analytical queries. Relevant studies published after 2015 were sourced from IEEE Xplore, ACM Digital Library, and ScienceDirect, focusing on empirical performance metrics. The findings highlight SQL's strength in structured data processing, achieving query speeds of up to 10ms for transactional workloads, while NoSQL excels in unstructured environments with 20–30% faster query speeds. HTAP systems combine features of SQL and NoSQL but face challenges in indexing strategies to optimize diverse workload types. The review concludes that database selection should be context-driven, optimizing for performance and scalability. The direction for future research will be the empirical testing of HTAP systems in real large-scale environments for the optimization of indexing strategies catering to both transactional and analytical demands.

Keywords: *schema optimization, indexing in large databases, SQL vs NoSQL, HTAP database performance*