



PROJECT REPORT
ANALYZING THE EFFECT OF INDEXING AND
CLUSTERING TECHNIQUES ON QUERY
PERFORMANCE IN E-COMMERCE TRANSACTION
DATABASES

ROSWITA KASIAMI GULO
22.K1.0041

Faculty of Computer Science
Soegijapranata Catholic University
2026

ABSTRACT

Query execution performance is a critical factor in database management, particularly in e-commerce transaction systems that continuously handle increasing data volumes. Without proper optimization techniques, operations such as data retrieval, updates, and deletions can take longer to execute and reduce overall system efficiency. This study aims to analyze the effect of index structure and clustering techniques on query execution performance in relational database systems. The dataset used in this study is the E-Commerce Sales Dataset, which consists of customer and transaction data integrated into a single dataset containing 9,744 records. Data processing and clustering were performed using the Python programming language with the pandas and scikit-learn libraries, while query performance testing was conducted on three database management systems: MySQL, PostgreSQL, and SQL Server. The experiment was conducted by comparing four test conditions: a baseline without optimization, using clustering, using indexing, and a combination of indexing and clustering. The index types implemented included B-Tree Indexes, Composite Indexes, Hash Indexes, and Non-Clustered Indexes. The results show that implementing indexing significantly improves performance across nearly all query operations. For example, for a DELETE operation, execution time is reduced from approximately 9 ms to 0.4122 ms, while for an UPDATE operation, execution time is reduced from 12.489 ms to 1.4002 ms under indexed conditions. Overall, these findings indicate that indexing is the most influential factor in improving query execution efficiency compared to clustering alone.

Keyword: indexing, clustering, query performance, e-commerce, relational database

