

Optimizing queries in distributed databases

Vu Thach An

Abstract: This study addresses the critical challenge of query optimization in distributed database systems, where data fragmentation and network latency significantly degrade performance. Traditional optimization techniques often fail to scale in heterogeneous environments due to the complexity of join ordering and data transfer costs. This research proposes a novel cost-based optimization model that integrates a dynamic programming algorithm with a heuristic pruning strategy to minimize total query execution time. The model evaluates both local processing costs and communication overhead across distributed nodes. Experimental results, conducted on a simulated multi-site database, demonstrate that the proposed approach reduces query response time by an average of 28% compared to conventional greedy algorithms, while maintaining comparable optimization overhead. Furthermore, the method shows robust performance under varying data distributions and network bandwidth conditions. These findings contribute to efficient data management in cloud-based and edge computing architectures, offering a scalable solution for real-time analytical workloads.

Keywords: Query optimization, distributed database, query processing, performance