A PARADIGM SHIFT IN DATABASE OPTIMIZATION: FROM INDICES TO AGGREGATES

Ryan C. LaBrie
Robert D. St. Louis
Lin Ye
Arizona State University

Abstract
Multidimensional databases and online analytical processing (OLAP) tools provide new ways for decision-makers to access data and retrieve information. This paper examines the differences between the optimization techniques that database designers need to consider when developing relational versus multidimensional data warehouses. The multidimensional data storage model allows for large numbers of aggregates to be stored in a very efficient and accessible manner. These aggregates make it possible not only to access information faster, but also at a lower cost in terms of CPU, I/O, and disk space utilization. This research-in-progress demonstrates those speed and cost savings using the TPC-H decision support benchmark. This publicly available dataset is used to contrast the use of indices in its relational implementation with the use of aggregates in its dimensional implementation. These contrasts show why database designers must shift from the index paradigm for relational databases to the aggregate paradigm for dimensional databases.

Keywords: Aggregates, data marts, indices, multidimensional databases, OLAP