

Abstract

Performance of a database is crucially dependent on its physical design. Current techniques, automated or otherwise, for physical design depend on the identification of a representative workload. However, in several emerging applications such techniques are inadequate as workload characteristics change rapidly over time. This is remarkably shown at the proxy cache of SkyQuery, a federation of Astronomy databases, which receives a continuously evolving workload. Using the proxy cache of the Skyquery federation as our case study, we present novel techniques for automated physical design of its database. We improve physical design of the proxy cache by vertical partitioning. Unlike prior workload-based vertical partitioning techniques that are offline, online partitioning algorithm adapts to workload changes incrementally and balances the performance benefits of physical design decisions with the cost of implementing these decisions. Our formulation includes a competitive algorithm based on task systems and an incremental algorithm based on association rules. The algorithms are general in that they do not make assumptions about the incoming workload or the underlying physical schema. Experiments with SkyQuery workloads show that incorporating online vertical partitioning in the proxy cache of SkyQuery databases provides a significant improvement in cache response time and efficiency. The algorithms incur a low optimization overhead compared with offline vertical partitioning techniques.