Automated Database Workload Characterization, Mapping, and Tuning through Machine Learning

Madeline MacDonald
University of Utah

UUCS-18-007

School of Computing
University of Utah
Salt Lake City, UT 84112 USA

10 December 2018

Abstract

Database management systems (DBMS) have numerous configuration settings, such as buffer or memory sizes, that have an impact on the database's performance. These settings require tuning to optimize performance, and the optimal values for each setting are highly dependent on the server's workload and hardware. The process of tuning a DBMS requires a highly skilled database administrator (DBA) to manually test different configurations, but this is an long and expensive process. Furthermore, the final tuning configuration is specific to the workload, DBMS version, and hardware, and if any of those factors change the tuning process must be redone.

OtterTune, a project by the Database Group at Carnegie Mellon University, solves this tuning problem by using machine learning techniques to automatically tune databases. OtterTune observes a new database workload, isolates the most important system metrics from that workload, and then maps the new workload to a similar previously tuned workload. It learns from previous tuning results for the similar workload and generates a new configuration recommendation automatically.

In this work, analyze OtterTune's architecture and approach to automating database tuning, specifically focusing on the workload characterization and mapping techniques. We then implement a lightweight version of OtterTune that automatically tunes PostgreSQL 9.6 databases. We utilize open source OtterTune code from Carnegie Mellon in this solution to ensure that the results are comparable. Finally, we test our implementation of OtterTune on a selected set of database benchmarks and analyze our results.