



Lin Ma

Self-Driving Database Management Systems: Forecasting, Modeling, and Planning

Wednesday, August 18, 2021 – 12:00 p.m. – REMOTE

Database management systems (DBMSs) are an important part of modern data-driven applications. However, they are notoriously difficult to deploy and administer because they have many aspects that one can change that affect their performance, including database physical design and system configuration. There are existing methods that recommend how to change these aspects of databases for an application. But most of them require humans to make final decisions on what changes to apply and when to apply them. Furthermore, these previous tuning methods either (1) require expensive exploratory testing, (2) are reactionary to the workload and can only solve problems after they occur, (3) focus only on improving one single aspect of the DBMS, or (4) do not provide explanations on their decisions. Thus, most DBMSs today still require onerous and costly human administration.

In this thesis, we present a novel architecture for a self-driving DBMS that enables automatic system management and removes the administration impediments. Our approach consists of three frameworks: (1) workload forecasting, (2) behavior modeling, and (3) action planning. The workload forecasting framework predicts the query arrival rates under varying database workload patterns using an ensemble of time-series forecasting models. The behavior modeling framework constructs fine-grained machine learning models that predict the runtime behavior of the DBMS. Lastly, the action planning framework generates a sequence of optimization actions based on these forecasted workload patterns and behavior model estimations. It uses receding horizon control and Monte Carlo tree search to approximate the complex optimization problem effectively.

Our forecasting-modeling-planning architecture enables an autonomous DBMS that proactively plans for optimization actions without expensive testing. It automatically applies the actions at proper times, holistically controls all system aspects, and provides explanations on its decisions.

Thesis Committee:
Andrew Pavlo, Chair
Ruslan Salakhutdinov
Greg Ganger
Christopher Ré, Stanford University