Query-based Workload Forecasting for Self-Driving DBMSs
Lin Ma, Dana Van Aken, Ahmed Hefny, Gustavo Mezerhane, Andrew Pavlo, Geoffrey J. Gordon

Motivation
- Database optimization requires experience and takes significant time.
- Self-driving DBMSs will reduce the complications and costs involved with deploying a database.
- Workload forecasting resilient to hardware and database system design is crucial for autonomous optimizations.

QB5000 Framework Pipeline

Pre-Processor
- Raw SQL
- Template
- Arrival Rate History
- 1 Hour Interval
- 3 Hour Interval

Clusterer
- Template Clusters
- KD Tree

Forecaster
- LR / KR
- RNN

QueryBot 5000: Overview
- Pre-Processor: Extract template for each query and record the arrival rate history for each template.
- Clusterer: Group similar templates together to reduce the modeling overhead.
- Forecaster: Predict the future arrival pattern for templates in each cluster at different horizons/interval.

Arrival Rate Forecasting
- Popular models in the literature
  - Linear Regression (LR), ARMA, FNN, Recurrent Neural Network (RNN), Kernel Regression (KR), PSRNN.
- ENSEMBLE: LR+RNN gives the best average accuracy.

Cluster SQL Templates on Arrival Rates
- Feature Options: Physical X Logical X Arrival Rate ✓
- Top 4 query templates within the largest cluster.

Example: Automatic Index Building
- Building indexes for the DBMS according to the workload forecasting in real time.
- Use the same index suggestion algorithm to build 20 indexes in total in all the settings.
- At the end, AUTO (using QB5000 in real time) achieves 25% better performance than STATIC (building all indexes at first using static workload sample).

Carnegie Mellon University