LlamaTune

Sample-Efficient DBMS Configuration Tuning

presented by Nikhil Yachareni
Introduction

- tuning DBMS
- large parameter spaces / hundreds of knobs
- domain knowledge
- sample efficiency

LlamaTune

- tuner design leveraging domain knowledge to improve sample efficiency of existing optimizers
Layout

- Introduction
  - Why a llama?
  - ML-based Tuning
  - DBMS Knob Tuning Procedure
  - Features + Contributions
  - LlamaTune Knob Tuning Procedure
  - LlamaTune Test Flow
  - Results
  - Authors
Background: ML-based tuning, knobs

2 predominant ways (but RL-based also considered)
- prior training + fine-tuning
- direct tuning (focus of this paper)
  - iteratively select configuration using optimizer
  - run workloads on them
  - exploration v/s exploitation
  - SMAC (https://github.com/automl/SMAC3)

Knobs
- discrete, continuous, categorical, hybrid knobs (special values)
DBMS Knob Tuning Procedure

Figure 1: Overview of DBMS Knob Tuning Procedure
Features + Contributions

Features

(1) Automated dimensionality Reduction
(2) Biased-sampling for special knob values
(3) Knob value bucketization for search space reduction

Contributions

(1) Shown benefit from tuning dimensionality-reduced important knobs
(2) Special knob value handling
(3) Large knob value range handling

Limitation

- Bucketizing entire search space may affect fine tuning on continuous knobs
Figure 8: LlamaTune: Tuning example that highlights the unified end-to-end pipeline.
LlamaTune Test Flow

- 3 Optimizers (2 BO-based, 1 RL-based)
- 6 popular OLTP workloads
- Tuning for optimizing throughput OR tail latency
- Porting to new DBMS version (~ 4 hours for full integration)
Results

Figure 9: Best throughput achieved by LlamaTune. Time-to-optimal also shown.\(^3\)

Figure 10: LlamaTune convergence gains vs. SMAC.

3 - Note that y-axis limits are chosen to improve readability of graphs. The only point below Y-axis minimum is the default configuration which is iteration 0.
## Authors

<table>
<thead>
<tr>
<th>Name</th>
<th>Institution</th>
<th>Email</th>
</tr>
</thead>
<tbody>
<tr>
<td>Konstantinos Kanellis</td>
<td>University of Wisconsin-Madison</td>
<td><a href="mailto:kkanellis@cs.wisc.edu">kkanellis@cs.wisc.edu</a></td>
</tr>
<tr>
<td>Cong Ding</td>
<td>University of Wisconsin-Madison</td>
<td><a href="mailto:congding@cs.wisc.edu">congding@cs.wisc.edu</a></td>
</tr>
<tr>
<td>Brian Kroth</td>
<td>Microsoft Gray Systems Lab</td>
<td><a href="mailto:bpkroth@microsoft.com">bpkroth@microsoft.com</a></td>
</tr>
<tr>
<td>Andreas Müller</td>
<td>Microsoft Gray Systems Lab</td>
<td><a href="mailto:amueller@microsoft.com">amueller@microsoft.com</a></td>
</tr>
<tr>
<td>Carlo Curino</td>
<td>Microsoft Gray Systems Lab</td>
<td><a href="mailto:ccurino@microsoft.com">ccurino@microsoft.com</a></td>
</tr>
<tr>
<td>Shivaram Venkataraman</td>
<td>University of Wisconsin-Madison</td>
<td><a href="mailto:shivaram@cs.wisc.edu">shivaram@cs.wisc.edu</a></td>
</tr>
</tbody>
</table>
Thank you!

Questions?