OTTERTUNE

Johannes Freischuetz

November 15th

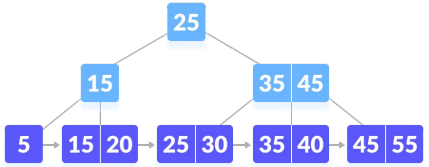
00. Background
01. The Academic Version
02. The Company
03. The Future of AutoTuning

00. Background

Optimizing Databases is complex and a “dark art”

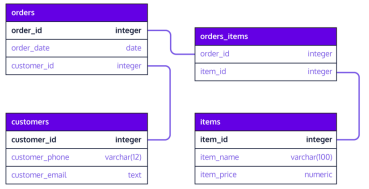
What things are tuneable?

Knobs



Indexes

Cloud Configuration



Schema Recommendations

00. Background

Occupational Employment and Wages, May 2022

15-1242 Database Administrators

Administer, test, and implement computer databases, applying knowledge of database management systems. Coordinate changes to computer databases. Identify, investigate, and resolve database performance issues, database capacity, and database scalability. May plan, coordinate, and implement security measures to safeguard computer databases. Excludes "Information Security Analysts" (15-1212) and "Database Architects" (15-1243).

National estimates for Database Administrators
Industry profile for Database Administrators
Geographic profile for Database Administrators

National estimates for Database Administrators:

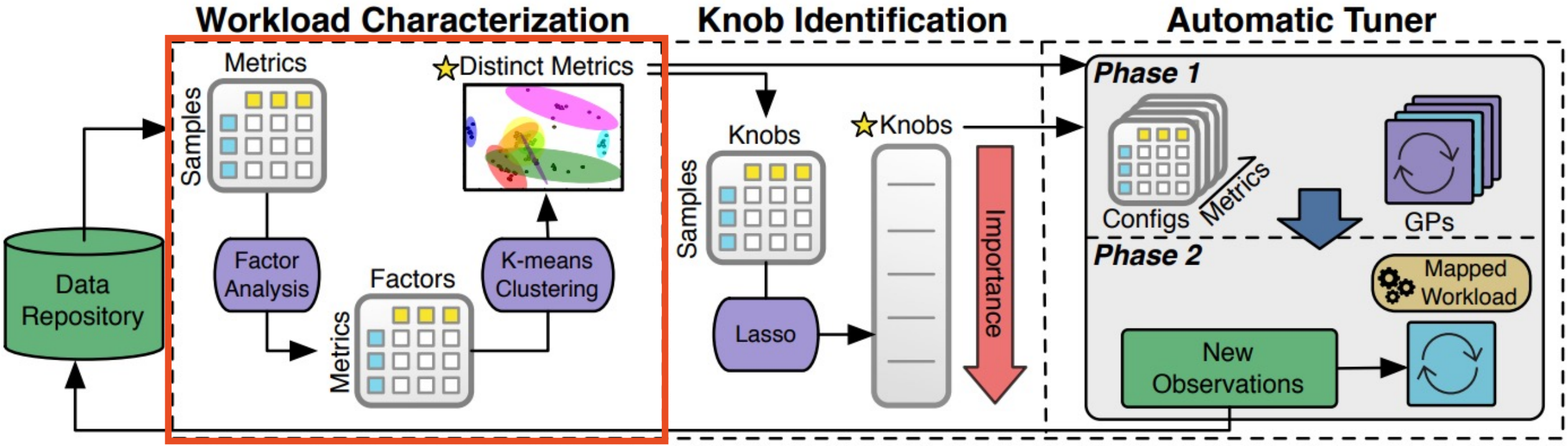
Employment estimate and mean wage estimates for Database Administrators:

Employment (1)	Employment RSE (3)	Mean hourly wage	Mean annual wage (2)	Wage RSE (3)
80,520	1.1 %	$ 49.29	$ 102,530	0.6 %

Percentile wage estimates for Database Administrators:

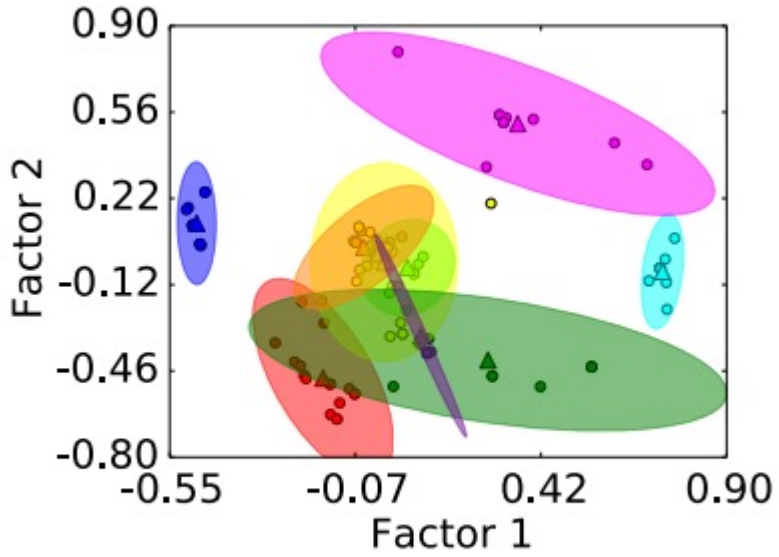
Percentile	10%	25%	50% (Median)	75%	90%
Hourly Wage	$ 25.49	$ 33.72	$ 48.03	$ 62.37	$ 73.98
Annual Wage (2)	$ 53,010	$ 70,140	$ 99,890	$ 129,730	$ 153,870

01. The Academic Version

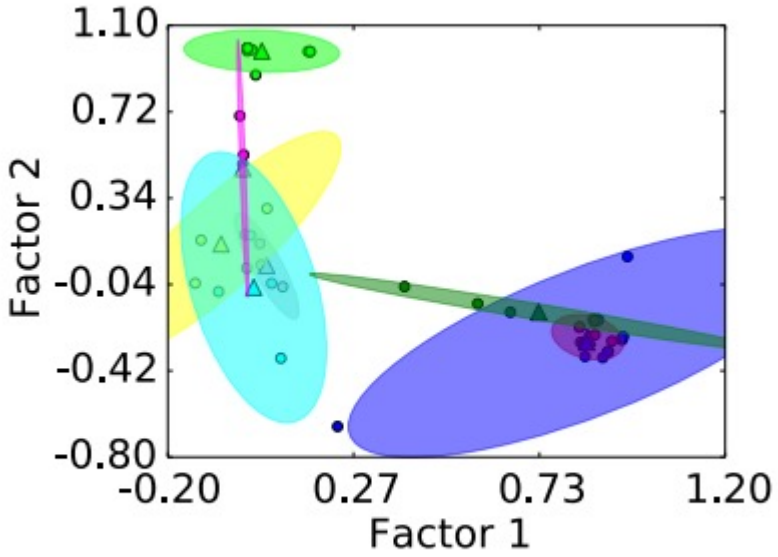


01. The Academic Version

Use factor analysis and clustering techniques for data reduction and to find distinguishing characteristics of the workloads

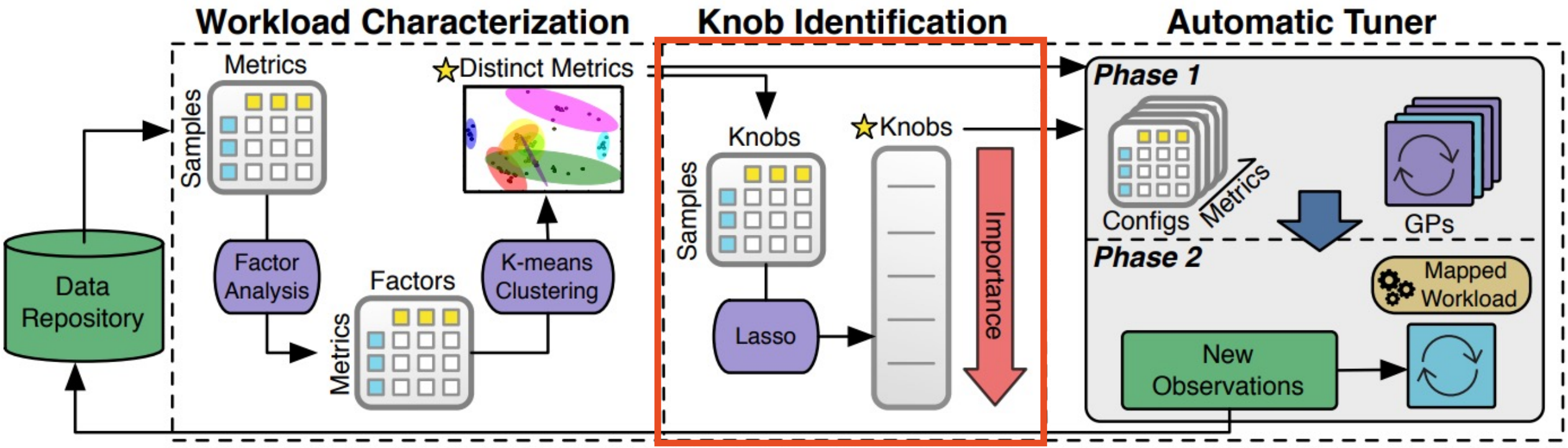


(a) MySQL (v5.6)



(b) Postgres (v9.3)

01. The Academic Version



01. The Academic Version

Use knob importance scores to determine which knobs to tune

In the original paper the use Lasso, modern systems use other metrics

LlamaTune: Sample-Efficient DBMS Configuration Tuning

Konstantinos Kanellis
University of Wisconsin-Madison
kkanellis@cs.wisc.edu

Cong Ding
University of Wisconsin-Madison
congding@cs.wisc.edu

Brian Kroth
Microsoft Gray Systems Lab
bpkroth@microsoft.com

Andreas Müller
Microsoft Gray Systems Lab
amueller@microsoft.com

Carlo Curino
Microsoft Gray Systems Lab
ccurino@microsoft.com

Shivaram Venkataraman
University of Wisconsin-Madison
shivaram@cs.wisc.edu

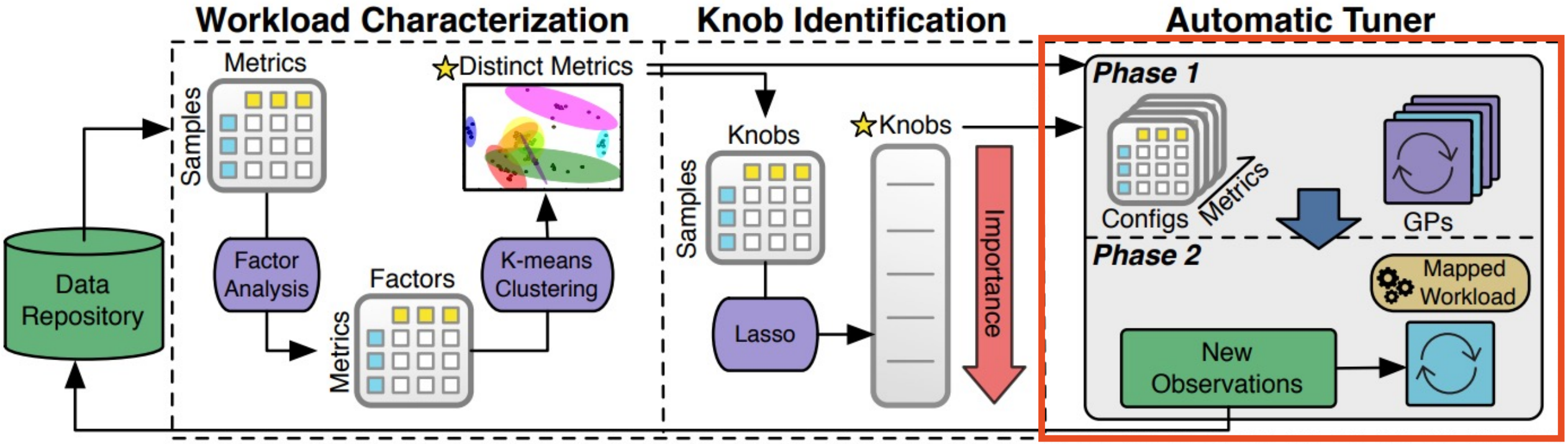
Aug 2022

ABSTRACT

Tuning a database system to achieve optimal performance on a given workload is a long-standing problem in the database community. A number of recent works have leveraged ML-based approaches to guide the sampling of large parameter spaces (hundreds of tuning knobs) in search for high performance configurations. Looking at Microsoft production services operating millions of

and the large number of configuration knobs [11, 32] makes it challenging to manually tune databases for high performance. As a result, a number of automated methods [3, 11, 14, 18, 33, 35, 38, 40, 41] for tuning have been proposed, with recent methods using machine learning (ML). ML-based methods have been shown to generalize to a number of workloads [39] and can find configurations that achieve up to 3-6× higher throughput (or lower latency) [18, 33, 38]

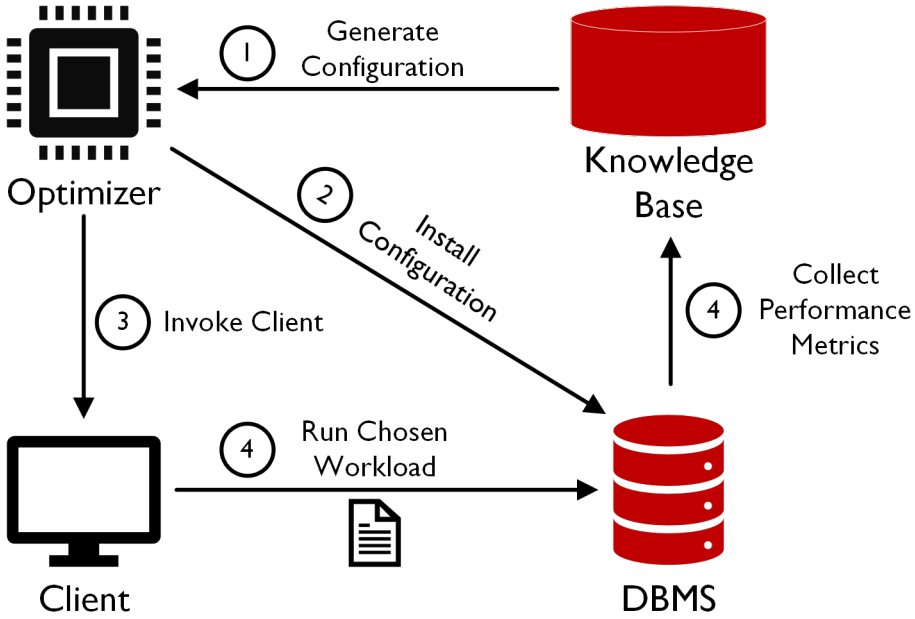
01. The Academic Version



01. The Academic Version

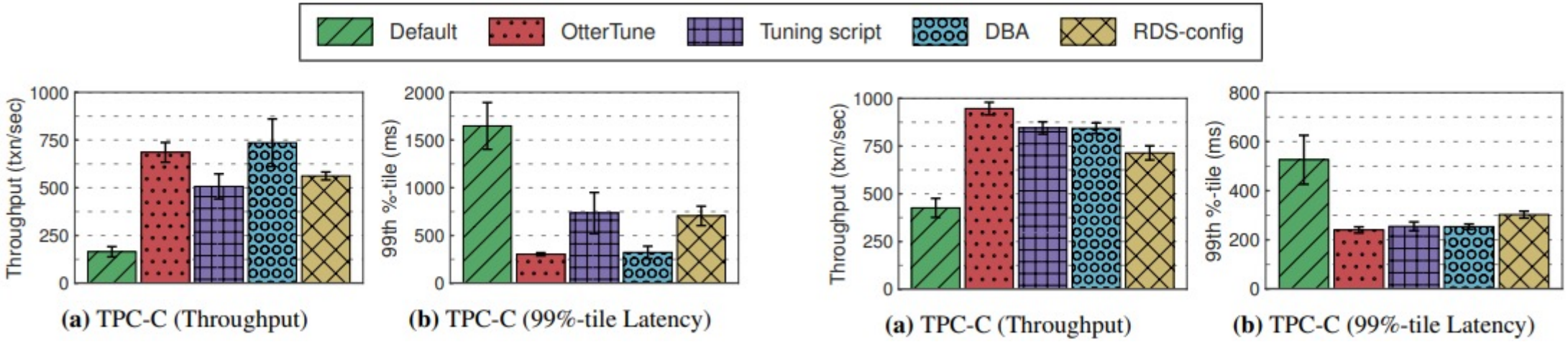
First does a workload mapping to bootstrap Gaussian Process Model with data from prior workload

Trades off exploration and exploitation in a very similar way to EGO



01. The Academic Version

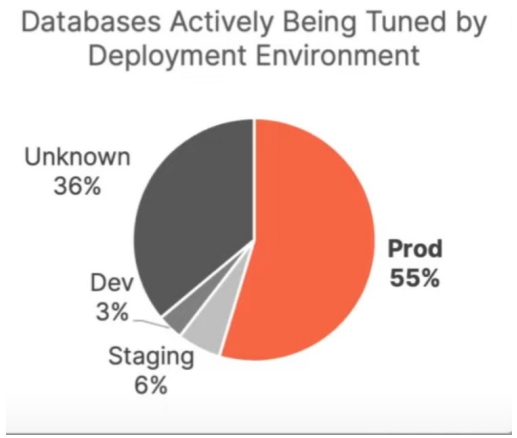
Performs significantly better than the default and almost as well as a Database Administrator



02. The Company

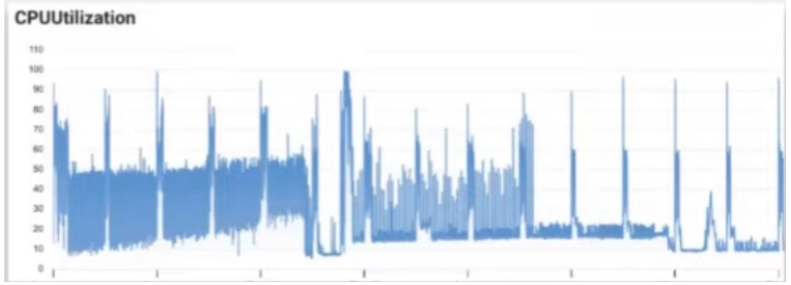
There are assumptions made in research that do not transfer to the real world

- Optimizations are run only one time, which assumes no workload skew
- DBAs have tools to deploy a copy of the production database



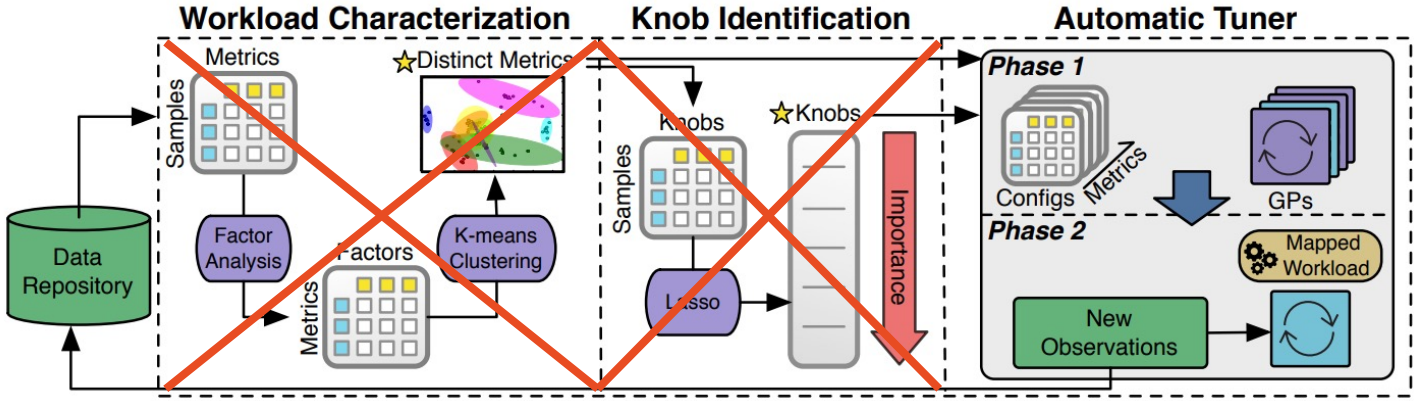
02. The Company

- Workloads can be represented with a short observation period
- A knowledge base of high-quality training data is available or practical to obtain



02. The Company

- We do we do in the real system to address this?
  - Incorporate Domain Knowledge in knob selection
  - Suggest 10 – 20 knobs to DBA and let them select which they want to tune
  - Speed up convergence by starting with recommendations from PGTune or MySQL Tuner
- They do not do Workload Characterization or Knob Identification anymore



RELATIONAL RIVERSIDE RUMBLE 2022

COMPETE FOR
$10,000 USD
CASH MONEY

HUMAN VS OTTERTUNE

00. The Future of Auto Tuning

- Sample Efficient Tuning Techniques
  - LlamaTune (VLDB ‘22)
  - Performance Roulette (NeurIPS MLforSystems ‘23)
- Workload Synthesis
  - Zhang et al. (VLDB ‘22)
- Starting and Stopping Criteria
- Techniques other than Bayesian Optimization
  - Qtune (VLDB ’22)
- Tuning Other Database Components
  - Cosine (VLDB ‘22)
  - Balsa (SIGMOD ‘22)