UNIVERSITY OF CAMBRIDGE
DEPARTMENT OF COMPUTER SCIENCE AND TECHNOLOGY

https://www.cst.cam.ac.uk/

Leveraging **I**nstance-**A**ware **I**ndex **A**dvising (IA2) with Reinforcement Learning for Diverse Workloads

Taiyi Wang (Taiyi.Wang@cl.cam.ac.uk)

Eiko Yoneki (eiko.yoneki@cl.cam.ac.uk)

University of Cambridge

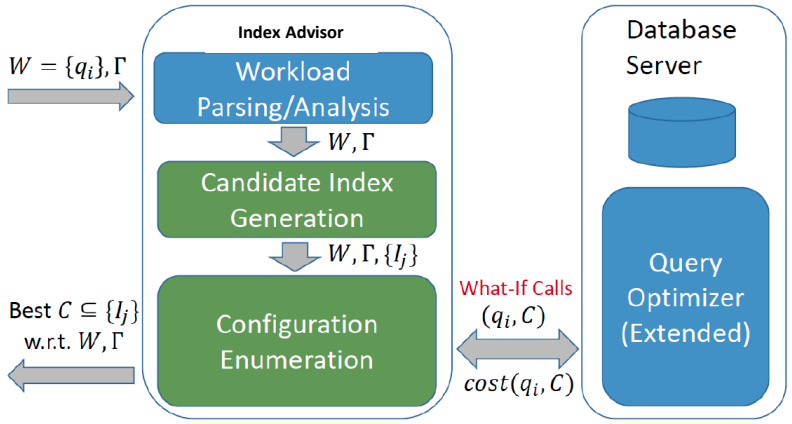
22/04/2024

4th EuroMLSys workshop- EuroSys'24, Athens, Greece

UNIVERSITY OF CAMBRIDGE

Motivation

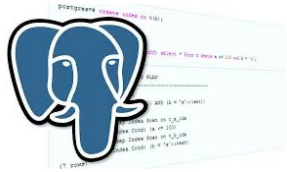
Multi-attribute Index Configuration for modern SQL-DB is not fully optimized*

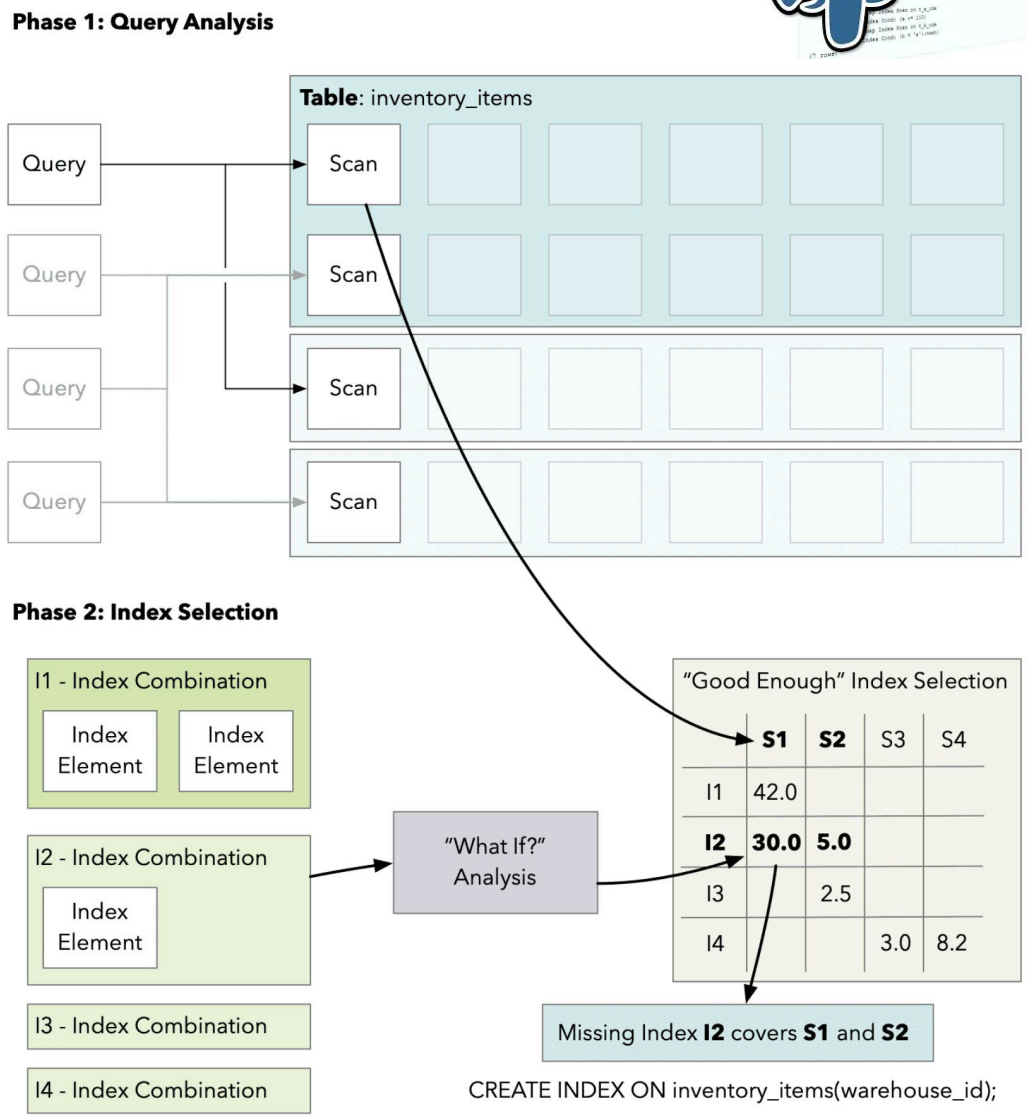


Index Advisor needs to

1. **Look into Queries/Workloads**
2. **Consider Budgets (Time/Storage)**
3. **Compare performances among candidates**

Cited from **pganalyze** https://pganalyze.com/





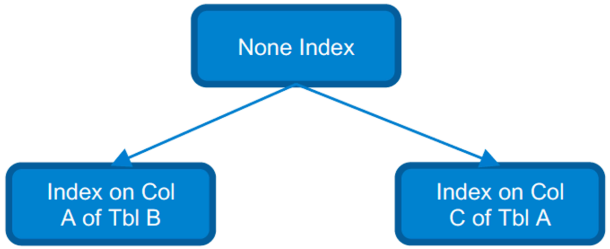
* Siddiqui, Tarique, and Wentao Wu. "ML-Powered Index Tuning: An Overview of Recent Progress and Open Challenges." *ACM SIGMOD* (2024)

UNIVERSITY OF CAMBRIDGE

Index Selection Problem is a Combinatorial Optimization (CO) Problem

Definition *Index Selection Problem (ISP): Given a workload W, a database D, a set of constraints C, and a set of candidate indexes $\mathbb{X}$, find an index configuration X such that $X^* = argmin_{X \subseteq \mathbb{X}} Cost(W, X)$ subject to all constraints in C.*

- How to find the best combinations for Multi-attributes Index?

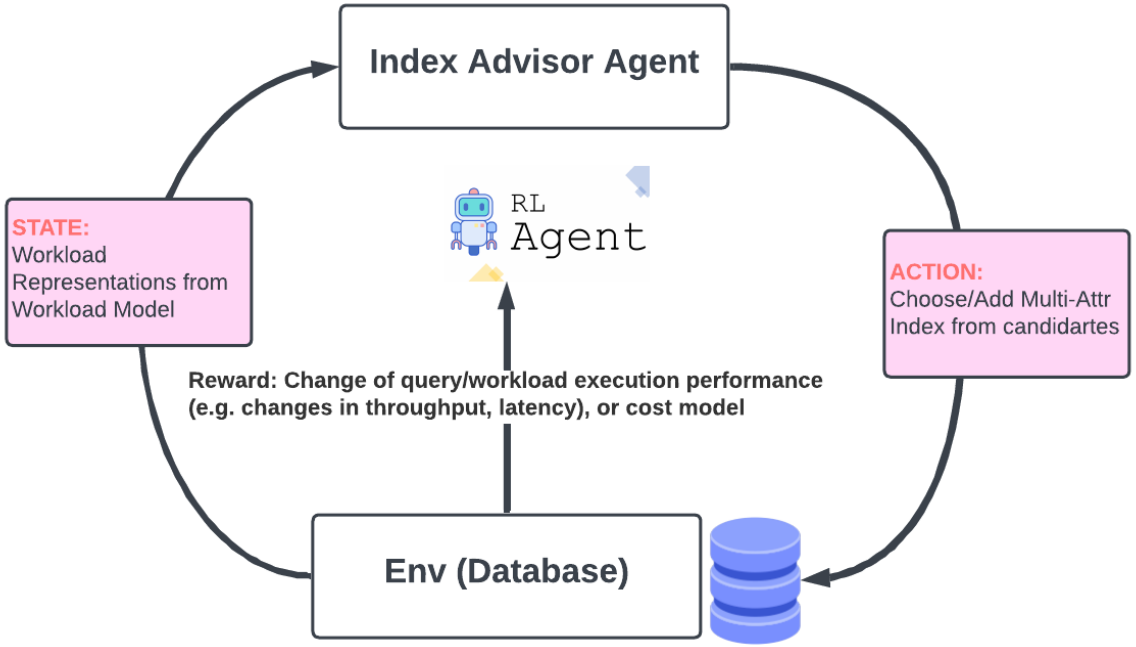


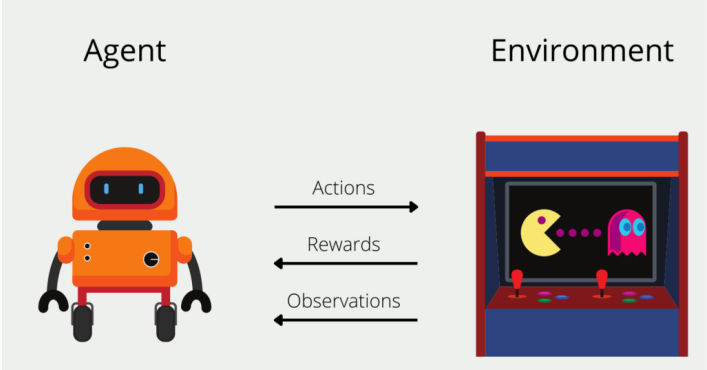
Deep Reinforcement Learning (DRL) is a good CO solver*

Learning Ability: DRL can generalize across different instances of a problem, providing robust solutions for different states visited

Scalability: DRL methods can scale to handle complex and large-scale problems by leveraging advanced neural network architectures

Efficient Exploration: DRL agents continuously improve their policies through interactions with the environment, leading to increasingly effective solutions over time.





* Dai, Hanjun, et al. “Learning combinatorial optimization algorithms over graphs.” NeurIPS (2017)

UNIVERSITY OF CAMBRIDGE

Action/Candidate Space Dilemma

- Index candidates will show an increase in the number of combinations, which is a huge challenge for RL training.

E.g. , TPCH workloads tested

- index width = 3 -> |A| = 639
- index width = 4 -> |A| = 1701

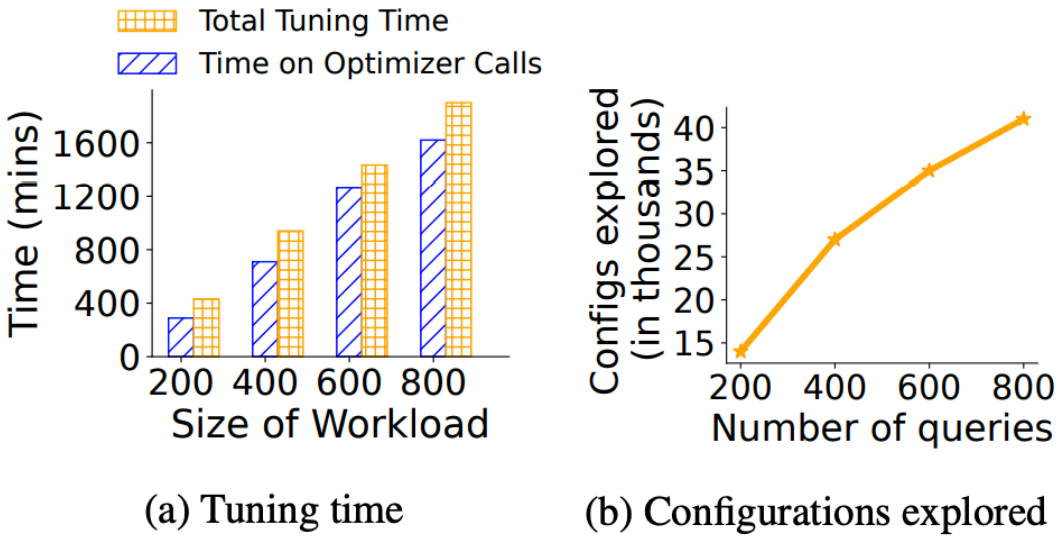


Figure 1

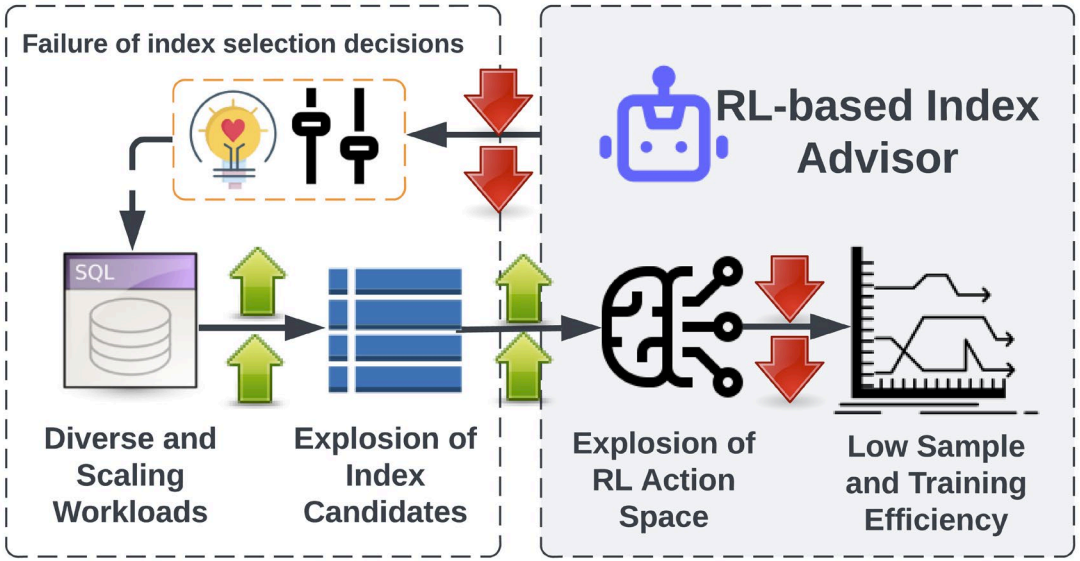


Figure 2

Figure 1: Siddiqui, Tarique, and Wentao Wu. "ML-Powered Index Tuning: An Overview of Recent Progress and Open Challenges." *ACM SIGMOD* (2024)
Figure 2: Wang et al. " Leveraging Instance-Aware Index Advising (IA2) with Reinforcement Learning for Diverse Workloads" EuroMLSys 2024

UNIVERSITY OF CAMBRIDGE

Existing Solutions dealing with Action Space

- Lan et al. CIKM'20

➔ Rule-based candidates' generation

Pros:

- Control and Efficiency: Allows easy control over the process with high training efficiency.

Cons:

- Limited Exploration: Risks missing out on potential candidates due to restrictive rules.
- Complex Rule Design: Creating comprehensive rules that cover all scenarios is challenging and can be resource-intensive.

- Kossmann et al. "SWIRL" EDBT'22

➔ Rule-based candidates masking

Pros:

- Enhanced Action Consideration: Generates all possible combinations and applies masking to consider a broader range of possible actions.

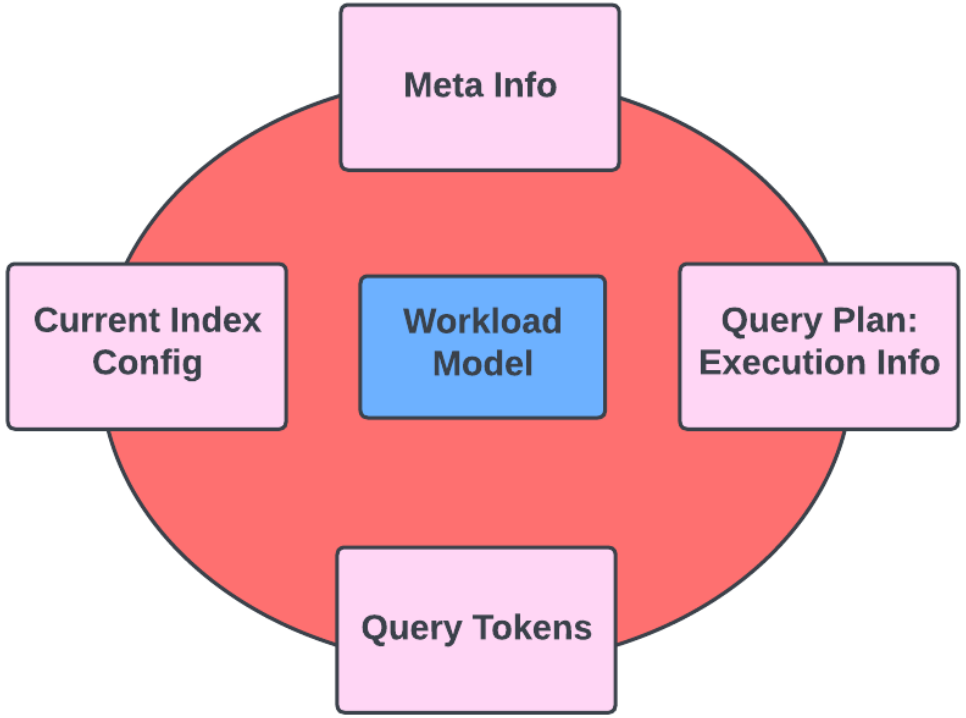
Cons:

- Variable Pruning Efficiency: Pruning efficiency varies with workload complexity, making it unreliable in some scenarios.
- Reduced Efficiency Under Complexity: Training efficiency decreases as workload complexity increases, potentially slowing down the learning process.

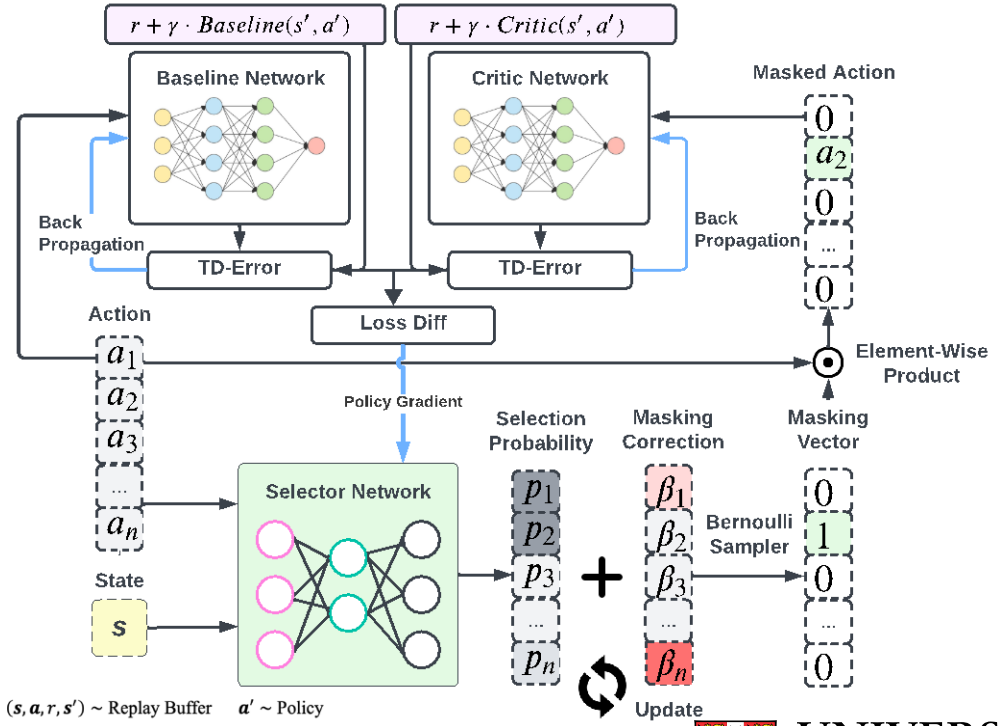
IA2 Core Design*

Capturing Strong Correlations between Valid Actions (Index Candidates) and States (Workload Representations)

- A generalizable workload model



- TD3-TD-SWAR: An adaptive action pruning block



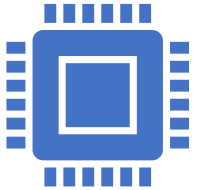
* Wang et al. ” Leveraging Instance-Aware Index Advising (IA2) with Reinforcement Learning for Diverse Workloads“ EuroMLSys 2024

UNIVERSITY OF CAMBRIDGE

Experimental Setup

Single PgSQL-DB, single thread

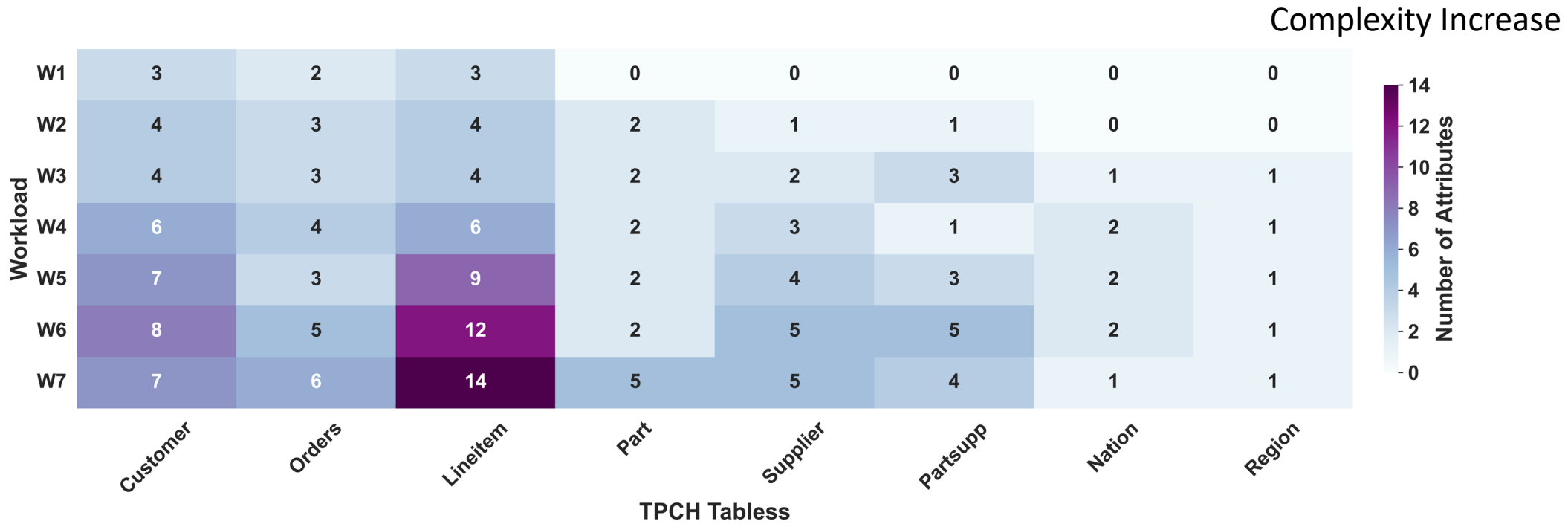
Benchmark Workloads: TPCH SF1



virtual machine equipped with a shared Nvidia Quadro RTX8000 GPU and
8 CPU cores, 64GB Memory.

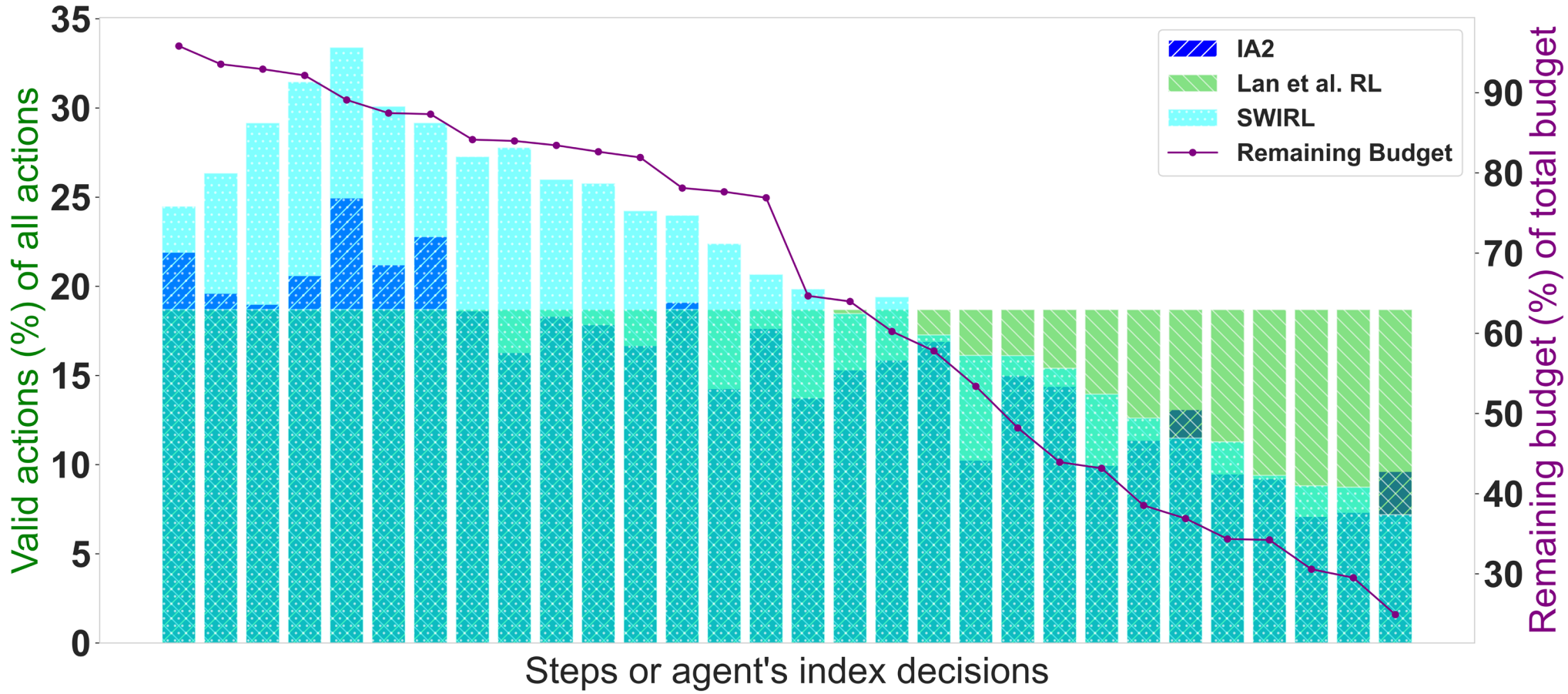
Although the experiments and the RL agent are not restricted to a specific type of index, the upcoming experiments will primarily use non-covering B-trees, which are the default index type in PostgreSQL.

Evaluated Workload (TPCH) Patterns



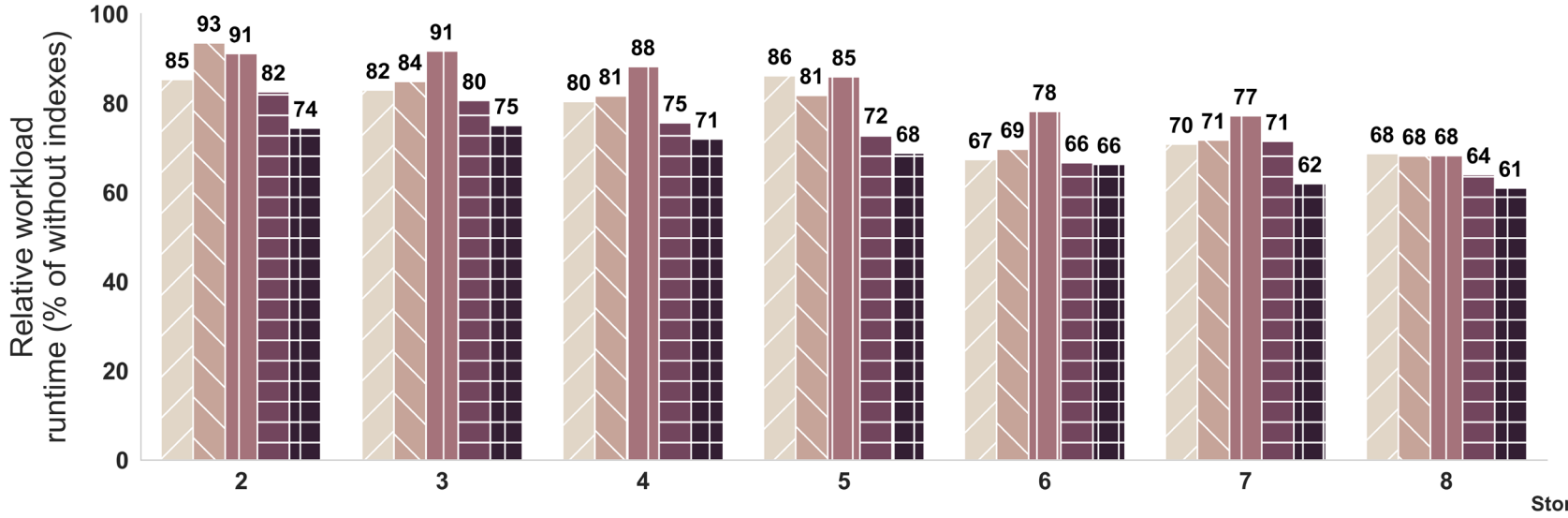
- Craft with TPCH-22 queries Templates
- Queries with heavy filters, Order by, Range conditions (e.g., q9, q6, q19, q18) are involved
- W1-W7: Increasing Complexity

Experiment – Action Pruning Efficiency

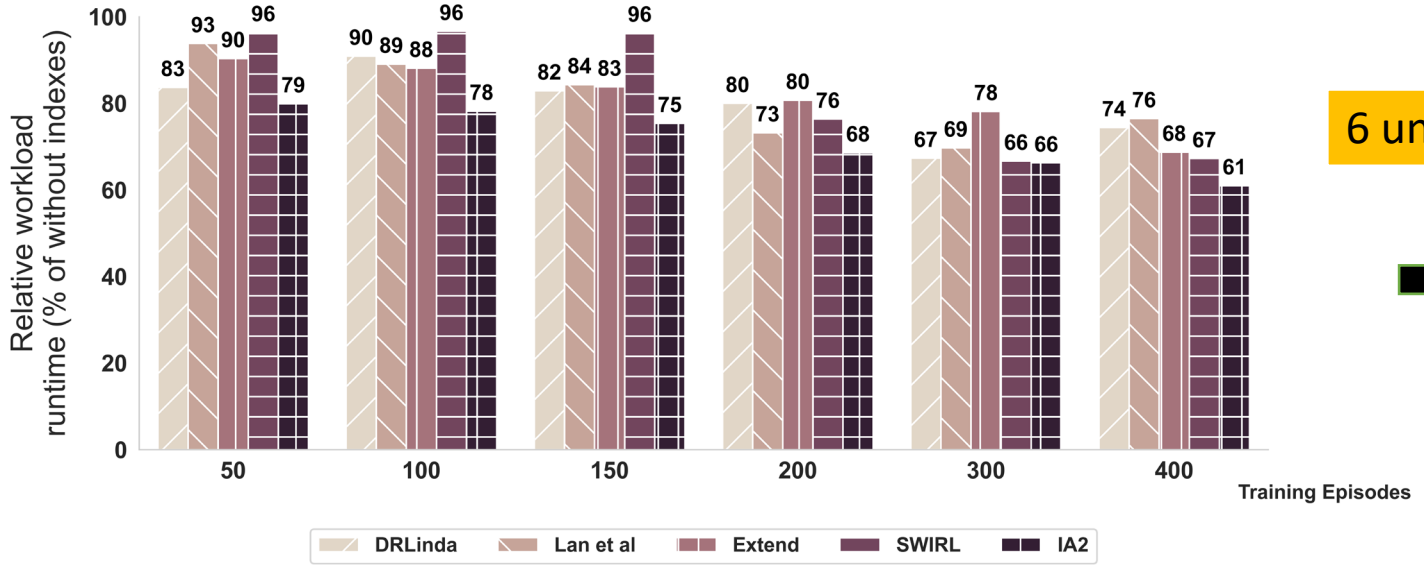


W6, budget = 8 unit, index width = 4, |A| = 1701

Experiment: Runtime Performance



1 unit = 128MB, W5, 300 epis



6 unit storage budget, W5

IA2 consistently outperforms other index selection methodologies by an average margin of 15-20%.

1. IA2 has high training efficiency
2. IA2 is a budget-aware advisor

Wang et al. ” Leveraging Instance-Aware Index Advising (IA2) with Reinforcement Learning for Diverse Workloads“ EuroMLSys 2024

UNIVERSITY OF CAMBRIDGE

Summary: Comparisons with Existing Works (RL-based Index Advisor)

Takeaways from IA2:

1.Rapid Training Efficiency
2.Advanced Workload Modeling
3.Effective Action Space Exploration

	DRLinda[1]	Lan et al.[2]	SWIRL [3]	IA2 [4]
Multi-attributes	No	Yes	Yes	Yes
Stop criterion	#Idx	#Idx,Storage	Storage	Storage
Workload representation	Yes	No	Yes	Yes
Gen. to new queries	++	-	+++	+++
Training difficulty	++	+	+++	+
Action Space	Raw	Rule-based gen.	Rule-based mask.	Rule-based gen. + Adaptive mask.

Extend [5]

- **Recursive strategy (Heuristic Method)**
- **Stop Criterion: Storage Budget**

[1] Sadri, Zahra, Le Gruenwald, and Eleazar Lead. “DRLindex: deep reinforcement learning index advisor for a cluster database.” ICDE 2020.
[2] Lan, Hai, Zhifeng Bao, and Yuwei Peng. "An index advisor using deep reinforcement learning." CIKM 2020.
[3] Kossmann, Jan, Alexander Kastius, and Rainer Schlosser. “SWIRL: Selection of Workload-aware Indexes using Reinforcement Learning.” EDBT 2022.
[4] Taiyi Wang, and Eiko Yoneki. ” Leveraging Instance-Aware Index Advising (IA2) with Reinforcement Learning for Diverse Workloads“ EuroMLSys 2024
[5] Rainer Schlosser, Jan Kossmann, and Martin Boissier. “Efficient scalable multi-attribute index selection using recursive strategies.” ICDE, 2019

Thanks!

UNIVERSITY OF CAMBRIDGE

Q & A

Email: Taiyi.Wang@cl.cam.ac.uk
Website: https://kevinwty0107.github.io/

UNIVERSITY OF CAMBRIDGE