

Введение. Настройка параметров является важным элементом оптимизации баз данных, позволяющим настроить значения параметров для оптимизации производительности базы данных. Однако в процессе настройки параметров существует несколько общих проблем. Во-первых, базы данных имеют множество параметров конфигурации, и человеку трудно определить параметры, изменение которых способно улучшить производительность базы данных. Во-вторых, большинство параметров имеют непрерывные значения, в пространстве которых эвристическая настройка параметров не сможет найти лучшие значения параметров в ограниченное время. В-третьих, настройка параметров становится все более сложной в связи с тем, что базы данных имеют различные сценарии использования, и число параметров растет с каждой версией программного обеспечения.

Однако в последнее время было предложено множество методов настройки параметров с использованием алгоритмов машинного обучения с подкреплением [1]. Основной идеей такого обучения является построение моделей, собирающие статистику обработки запросов базы данных, на основе которых производится дальнейшее обучение моделей.

В данном исследовании представлен обзор существующих моделей для настройки параметров баз данных, а также их сравнительный анализ на примере оптимизации базы данных PostgreSQL с предметной областью в виде онлайн-магазина. Настройка параметров включает в себя выбор параметров базы данных, построение моделей машинного обучения, а также передачу данных между моделью и базой данных.

Основная часть. В данной работе проводится исследование оптимизации параметров базы данных PostgreSQL версии 16.1.1 для онлайн-магазина. Пайплайн исследования включает следующие этапы: сначала происходит отбор настроек из общего числа (параметров порядка 360), затем проводится обучение различных моделей машинного обучения, таких как Deep Q-Network, Deep Deterministic Policy Gradient [2], Trust Region Policy Optimization [3] и других, с целью предложить оптимальные значения параметров. Основная метрика производительности, используемая для оценки эффективности моделей, это количество запросов в час. Также в качестве оценки сверху являются ресурсы информационной системы.

На этапе исследования особое внимание уделяется тому, что все необходимые индексы уже добавлены, а запросы оптимизированы для достижения наилучшей производительности. В работе идет фокус исключительно на настройке параметров базы данных с помощью различных моделей машинного обучения. Целью работы является поиск оптимальных значений параметров, способных максимизировать производительность базы данных в контексте работы онлайн-магазина, а также выявлению наилучшей модели, примененной к вышеуказанной предметной области.

Выводы. Произведен анализ моделей для оптимизации параметров базы данных и произведено качественное сравнение моделей по общей для всех моделей метрике.

Список использованных источников:

1. *Xinyang Zhao, Xuanhe Zhou, Guoliang Li* Automatic Database Knob Tuning: A survey. IEEE Transactions on Knowledge and Data Engineering, 2023, vol. 35, pp 12470-12490 – URL: <https://dbgroup.cs.tsinghua.edu.cn/ligl/papers/tuning-survey.pdf> (data access – 2 January 2024).
2. *Ji Zhang, Yu Liu, Ke Zhou, Guoliang Li, Zhili Xiao, Bin Cheng, Jiashu Xing, Yangtao Wang, Tianheng Cheng, Li Liu, Minwei Ran, Zekang Li* An End-to-End Automatic Cloud Database Tuning System Using Deep Reinforcement Learning. SIGMOD Conference, 2019, pp 415-

- 432 – URL: <https://dbgroup.cs.tsinghua.edu.cn/ligl/papers/sigmod19-cdbtune> (data access – 2 January 2024)
3. *John Schulman, Sergey Levine, Philipp Moritz, Michael I. Jordan, Pieter Abbeel*. Trust Region Policy Optimization. Proceedings of the 32nd International Conference on Machine Learning, 2015, pp – 1889-1897 – URL: <https://arxiv.org/pdf/1502.05477.pdf> (data access – 2 January 2024)