

JIA YUAN

Email : woodyuan@tencent.com — Tel : (+86) 17827063225

EDUCATION

SUN YAT-SEN UNIVERSITY, GuangZhou, China
Bachelor of Computer Science and Engineering

Sep. 2018 - Jun. 2022
GPA: 3.1/4

Major Course: Advanced Algebra; Discrete Mathematics; Probability and Statistics; Principles of Computer Organization; Data Structures and Algorithms; Principles and Operations of Supercomputers; Principles of Operating Systems; Computer Networks; Distributed Systems; Foundation Level Programming for High-Performance Computing; Artificial Intelligence; Natural Language Processing; Database Systems; Multi-Core Program Design and Practice; Computer Vision; Principles of Compilers

WORKING/RESEARCH EXPERIENCE

Tencent

ShenZhen, China

Cloud Database Research and Development Department — **AI for Database Project Team**

Position: Industry Research Intern

Oct. 2021 - Apr. 2022

Query Optimization - Cardinality Estimation

- Cardinality estimation is crucial for the database optimizer to generate execution plans, especially for index selection. It is a fundamental yet long-standing problem in the field of query optimization that has not been fully resolved.
- I conducted thorough research on this issue in both industry and academia. Finally, I proposed a method of using TOP K frequencies instead of sampling to build histograms for cardinality estimation. Test results have shown that this method achieves 40% higher accuracy in cardinality estimation compared to the original histogram approach in MySQL. Statistic the top K frequency for massive data may be very memory intensive. Based on the SpaceSaving algorithm, I propose a more optimized data structure to achieve memory controllability and time complexity of $O(n)$ for efficient streaming processing of large amounts of data.

Anomaly Detection and Root Cause Analysis

- For large-scale database clusters in the cloud, relying on manual discovery and analysis of anomalies and root causes is not feasible. It is crucial to have the capability of automated anomaly detection and root cause analysis through the combination of machine learning algorithms and DBA expertise.
- First phase of the project is 7x24 hour anomaly detection for all cloud instances. I implemented a demo version of the anomaly detection framework and tested various temporal data anomaly detection algorithms. Second phase of the project is SQL level root cause analysis, and I am also one of the main participants.

Position: Full Time Developer

Jul. 2022 - Present

cdbtune Project Phase 2

- The Tencent Cloud Database AI for Database project team has published two research papers on database self-tuning at SIGMOD conference: “An End-to-End Automatic Cloud Database Tuning System Using Deep Reinforcement Learning” and “HUNTER: An Online Cloud Database Hybrid Tuning System for Personalized Requirements”. These papers focus on the optimization of database performance through advanced techniques.
- The productized version of these research findings is known as cdbtune Phase 1. In cdbtune Phase 2, which I am responsible for, I have further enhanced the system’s optimization capabilities. Specifically, I have upgraded the tuning process from being based on sysbench/TPCC benchmarking to utilizing real user workload for more accurate and personalized performance tuning.

Workload Generation

- Workload generation is defined as generating a representative workload from a given load. However, current research in both academia and industry primarily focuses on ensuring relative sequential consistency of transaction execution, thereby overlooking conflicts between loads under concurrency. I have collaborated with Dr. Baoqing Cai, the first author of the paper “HUNTER: An Online Cloud Database Hybrid Tuning System for Personalized Requirements” to jointly develop a workload generation project. Currently, the research paper of this project is in the process of being submitted. Additionally, I have taken the lead in commercializing the workload generation, and its benchmarking capability has already been utilized in the cdbtune Phase 2 project. The load replay capability can achieve over 90% fidelity to the original load, and it is now widely used for internal anomaly diagnosis and commercial proof-of-concept (POC).

SKILLS

- **management and development capabilities of complex systems** — impressive engineering capabilities
- **integrate machine learning (ML) with systems** — extensive expertise in AI4DB
- **research ability** — capable of bridging the gap between academia and industry effectively
- **TOEFL** — in progress