Guanduo Chen

gdchen22@m.fudan.edu.cn | wooooooow-pro.github.io | wooooooow-pro | (86)18019204815

Education

Master of Science in Computer Science and Technology, Fudan University September 2022 –- July 2025 (Expected)

- Research Interests: MLSys, AI4DB

B.Eng in Computer Science and Technology, Fudan University September 2018 –- July 2022

Work Experience

ByteDance, Ltd. December 2023 -–April 2024

Database System R&D Intern, ByteHouse Runtime Team Shanghai, China

- **Geographic Data Aggregation Query Benchmark** Constructed a benchmark using the NYC Taxi dataset to compare the performance of mainstream databases (StarRocks, ClickHouse, PostGIS, DuckDB) against ByteHouse-CE for geographic data aggregation queries.
- **CNCH Geographic Data Type Support** Enabled ByteHouse-CNCH (Cloud Native ClickHouse) to support geographic data types using geos.
- **CNCH Geographic Primary Key Index Design and Implementation** Added support for geographic primary key indexes in CNCH and designed a multi-level cache (disk cache, memory cache) to accelerate index queries, reducing query latency by nearly 50% compared to native ClickHouse.

Research Experience

Computation Efficient LoRA March 2024 –- Present

Advisor: Prof. Binhang Yuan HKUST Relaxed System Lab

- Analyzed the computational bottlenecks in LoRA training and attempted to accelerate LoRA training by modifying the backpropagation computation method.

Learned Range Filter [Code] [Paper] February 2023 –December 2023

Advisor: Prof. Siqiang Luo NTU Data Management Research Group

- **Oasis** is a Learned Range Filter, which segments the key space into non-overlapping intervals and maps the data into a bitmap using a linear model-simulated CDF as the hash function. The filter utilizes block-based Elias-Fano compression to reduce space overhead without sacrificing query efficiency.
- **Oasis+** is a hybrid range filter that combines learning-based and hash-based methods to enhance filter applicability and robustness across various workloads.
- **Integrated SOTA filters into RocksDB** and tested average query response times and FPR under read-only workloads, achieving up to $6.2\times$ improvement in performance compared to existing range filters.

Text2SQL [Code] [Paper] July 2021 -–February 2022

Advisor: Prof. Xiaoyang Sean Wang Fudan DASLab

- Developed GAR (Generate-And-Rank), a Text2SQL framework that, given a set of sample SQL queries for a database, uses a parser to split the SQL into units. The generator recombines these units into a SQL set, which is then translated into text by a SQL2Text model. Finally, the retrieve strategy identifies the SQL matching the user's query.
- Constructed a complex Text2SQL benchmark with self-join semantics and analyzed the performance differences between GAR and end-to-end Text2SQL models based on this benchmark.

Projects

cmu-db/bustub January 2023 –- February 2023

- Implemented a K-LRU-based buffer pool management system.
- Designed and implemented a concurrent query-supporting extensible hash index and an optimistic-lock-based B+ Tree.
- Developed a basic query planner and rule-based query optimizer.

naive MIPS-CPU April 2021 –June 2021

- Implemented a single-core simulated CPU based on the MIPS architecture, supporting basic instruction sets and a simple pipeline using SystemVerilog.
- Added basic I/O and simple dynamic branch prediction functionality, and deployed the simulated CPU to hardware.

MIT6.828/xv6

- Developed a basic virtual memory management system, including physical address allocation and reclamation, and page table management.
- Designed the system Trap Frame and implemented system interrupts and system calls based on it.

Publication

1. *Oasis: An Optimal Disjoint Segmented Learned Range Filter* [VLDB 24]
   Guanduo Chen, Zhenying He, Meng Li, Siqiang Luo
2. *GAR: A Generate-and-Rank Approach for Natural Language to SQL Translation* [ICDE 23]
   Y Fan, Z He, T Ren, D Guo, L Chen, R Zhu, **G Chen**, Y Jing, K Zhang, XS Wang