

WENLONG WANG

+1(612)-978-1482

wang9467@umn.edu

EDUCATION

Ph.D. in Computer Science and Engineering

University of Minnesota, Twin Cities. Supervised by Professor David Hung-Chang Du.

Sept. 2019 - May 2026(Expected)

GPA: 3.76/4.0

B.S. in Computer Science and Engineering

Hong Kong University of Science and Technology.

Sept. 2015 - May 2019

First Honor. GPA: 3.67/4.3

WORKING EXPERIENCE

Research Intern | Futurewei Storage Lab, Santa Clara, CA

June 2022 - Aug 2022

Mentors: Dr. Chun Liu and Dr. Nelson Liao

- Researched updatable Learned Indexes, identifying performance impacts of order maintenance in leaf nodes and resolving slow performance in Single-Producer-Multi-Consumer (SPMC) situations.
- Designed and implemented a bucketed approach for leaf nodes, improving efficiency by up to 2.05x over existing works and laying the groundwork for a future publication.
- Extended the established framework of Learned Indexes to the string key index domain, broadening its applicability.

Research Intern | Alibaba Group (U.S.) Inc, Sunnyvale, CA

June 2021 - Aug 2021

Mentor: Dr. Sheng Qiu

- Researched optimizing the Flash Translation Layer (FTL) mapping on open-channel SSD during garbage collection in log-structured files (e.g., blob files in RocksDB) to reduce data reorganization overhead in physical space.
- Constructed the open-channel SSD benchmarking environment using FEMU and benchmarked the initial performance using BlobDB(RocksDB), which was used to guide future optimizations.

RESEARCH INTEREST

Distributed/Disaggregated cloud infrastructure, Large-scale KV stores, Learned Index, System and storage optimization, Machine learning systems, LLM serving optimization, Database systems, Emerging storage technologies (NVM, ZNS, hybrid SSD)

SELECTED PUBLICATIONS AND RESEARCH PROJECTS

LearnedKV: Integrating LSM and Learned Index for Superior Performance on SSD

- Designed and implemented a novel, large-scaled key-value store (over 4K lines of C++, excluding RocksDB) that seamlessly integrates an LSM tree with a Learned Index on storage, including a non-blocking conversion mechanism.
- Benchmarked our implementation with YCSB workloads and the results show that our work outperforms state-of-the-art solutions by up to 2.69x in read throughput and 1.43x in write throughput.
- **Publication (First author):** Preprint: <https://arxiv.org/pdf/2406.18892>. Full paper to be submitted to VLDB.

BLI: a High-Performance Bucket Learned Index with Concurrency Support

- Designed and implemented an updatable in-memory learned index that adopts a "globally sorted, locally unsorted" approach by replacing linear sorted arrays with unsorted Buckets and optimized tree height (with 6K lines of C++).
- Benchmarked our implementation with SOSD workloads and the results show that our work achieves up to 4.42x better throughput than state-of-the-art learned indexes, with even more gains under multi-threaded conditions.
- **Publication (Second author):** Selected in 2023 UMN CS&E Research Showcase. Full paper to be submitted to VLDB.

Optimizing Multi-GPU Data-Parallelism LLM Serving with Prefix Sharing

- Enhanced LLM serving efficiency by extending shared-prefix caching from single-GPU to multi-GPU environments.
- Designed and implemented algorithms for optimal prefix distribution and load balancing across GPU clusters, improving throughput and reducing latency for concurrent user requests in data-parallel LLM inference.

Zero-Copy Shard Migration in Disaggregated LSM-KVS

- Engineered a mechanism in disaggregated infrastructure to ingest SST files from RocksDB without physical copy overhead during shard migration and designed a cost-effect analyzer to selectively perform synchronous/asynchronous compaction.

VL-DNA: Enhancing DNA Storage Capacity with Variable Payload (Strand) Lengths

- Proposed and implemented a new scheme that used variable strand length to split primer-payload collisions and further enhanced the capacity of single tube DNA storage ranging from 18.27% to 19x (with 7K lines of C++).
- **Publication (Second author):** Preprint: <https://arxiv.org/pdf/2403.14204>. Selected in DAC 2023 WIP poster session.

ACADEMIC SERVICES AND AWARDS

- OSDI/ATC 2024 Artifact Evaluation Committee (AEC), MSST 2024 Reviewer, OSDI/ATC 2024 Student Grant.

SKILLS

- Programming Languages: C/C++, Python, JavaScript, HTML, Matlab
- Others: RocksDB, LevelDB, FEMU, HDFS, vLLM, Ray, Linux, Git, Data Structures and Algorithms, OS, Database