Ruihong Wang

Personal Page: https://ruihong123.github.io/ wang4996@purdue.edu
Github : https://github.com/ruihong123 (979) 777-6340

Personal Statement

I am a PhD candidate in CS department at Purdue University, co-advised by Prof. Walid G. Aref and Jianguo Wang. My research focuses on the cloud native database design. I would like to take an internship related to cloud computing or database design in 2025.

EDUCATION

Purdue University, West Lafayyete, IN
PhD, Computer Science
expected Dec 2025 GPA: 3.75/4.00

Texas A&M University, College Station, TX
Master of Science, Computer Engineering
Aug 2020 GPA: 4.00/4.00

Jilin University, Jilin China
Bachelor of Engineering , Electrical Engineering and Automation
June, 2018 GPA: 3.33/4.00

EXPERIENCE

Research Assistant **May 2022 - Fall 2023**
Purdue University IN, West Lafayette
Advised by Prof. Jianguo Wang, Prof. Walid G. Aref.

Research Intern **May 2023 - August 2023**
Alibaba US WA, Bellevue
Worked in the research group in PolarDB, responsible for the benchmarking and development of a level-2 buffer pool in PolarDB.

Teaching Assistant **Aug 2020 - May 2022**
Purdue University IN, West Lafayyete
Experienced in CS 348 Information System and CS 505 Distributed System.

Research Assistant **May 2019 - August 2020**
Texas A&M Transportation Institute TX, College Station
Worked in the road safety group, responsible for the data analysis.

REASEARCH

Cache Coherence Protocol over RDMA-Enabled Memory Disaggregation
Purdue University May 2022 - Now

- Implemented an innovative cache coherence protocol over RDMA-enabled disaggregated memory. The protocol is processed in a distributed manner by one-sided RDMA, making the memory nodes compute-free.
- Instantiate the protocol into an access engine over disaggregated memory, on which the traditional index design and concurrency control algorithm perform correctly.
- Optimized the performance of our protocol, and figured out the ideal workload for our access engine.
- Conduct a thorough benchmark study. The results show our protocol outperforms traditional 2-sided protocols (e.g., GAM) significantly in magnitudes.

Level-2 Caching Framework in PolarDB
Alibaba-US May 2023 - August 2023

- Contribute to the codebase of the secondary caching framework in PolarDB. This framework leverages those unutilized resources in the cloud (e.g., SSD, persistent memory and disaggregated memory) to improve cache hit ratio of queries.

- Benchmark the level-2 caching framework in PolarDB by sysbench and draft an academic paper.

LSM-Based Index for RDMA-Enabled Memory Disaggregation
Purdue University Aug 2020 - Now

- Implemented Log Structured Merge tree over disaggregated memory, applying architecture-specific optimizations (light-weighted in-memory write, byte-addressable queries, adaptive near-data compactions) to it.
- Brought up a method to do transactional-consistent checkpoints efficiently for in-memory LSM tree.
- Implemented a file system over the disaggregated memory and migrated the Rocks DB onto that FS.
- Compared dLSM with the state of art disaggregated b-tree, showing dLSM's high performance with the write-intensive workload.

Vehicle Category Classification based on the GPS Trajectory Data
Texas A&M University May 2019 - Aug 2020

- Applied Spark to deal with GPS trajectory points data.
- Analyzed the association rules based on the Maryland waypoint data and block-group attributes.
- Applied varying algorithms to the vehicle category classification and compared the performances.

PUBLICATIONS

Ruihong Wang, Jianguo Wang, Stratos Idreos, M. Tamer Özsu, Walid G. Aref. The Case for Distributed Shared-Memory Databases with RDMA-Enabled Memory Disaggregation. Proceedings of Very Large Data Bases Conference (VLDB), Volume 16, Issues 1, Pages 15-22, 2023.

Ruihong Wang, Jianguo Wang, Prishita Kadam, M. Tamer zsu, Walid G. Aref. dLSM: An LSM-Based Index for Memory Disaggregation. Proceedings of International Conference on Data Engineering (ICDE), pages 2835-2849, 2023.

Ruihong Wang, Chuqing Gao, Jianguo Wang, Prishita Kadam, M. Tamer zsu, Walid G. Aref. Optimizing LSM-based Indexes for Disaggregated Memory. Very Large Data Base Journal (VLDBJ), 2024.

AWARDS

Research assistant supported by Ross-Lynn Research Scholar Fund (2022-2023)

TECHNICAL SKILLS

Programming Languages : Python, C++, Java, C, Ruby, Assembly, HTML, R, Latex, SQL, shell.
Software/Framework : Pytorch, TensorFlow, Gephi, ibverbs, Qgis, Pyspark, make, cmake.