# Hongbin Zhong

+86 18379727287 rj986215159@outlook.com
github:https://github.com/rjzhb

## EDUCATION

**Georgia Institute of Technology** Aug 2024 - Jun 2029
Ph.D in Computer Science Atlanta

**Northeastern University** 985 Sep 2020 - Jul 2024
Bachelor in Computer Science China

## PUBLICATION

**[SIGMOD 2024]** PECJ: Stream Window Join on Disorder Data Streams with Proactive Error Compensation.
Xianzhi Zeng*, Shuhao Zhang(Corresponding author), **Hongbin Zhong**, Hao Zhang, Mian Lu, zhao zheng, Yuqiang Chen

**[code]** **[paper]**

## RESEARCH EXPERIENCE

**Columbia University** Jul 2023 - Nov 2023
Research Assistant New York City, NY
Adviser: ***Eugene Wu***

- *FADE Project - Query Explanation (Answering "Why" Made Fast)*
  - **Using Sparse Matrix**: explore and implement how interventions evaluation can be adopted to work with sparse arrays without compromising performance
  - **Avoiding Reading And Writing To Disk:** Optimized the generation of Interventions, significantly reducing disk I/O and allowing most of the required data for the evaluation phase to be directly accessed within the CPP file.
  - **Apply SIMD, MultiThreading Optimizations:** Apply SIMD and MultiThreading to the Evaluation loops of Sparse Matrix data structure.
  - **Benchmarks And Experiments:** Finished numerous benchmarks and completed a significant number of experiments.

**Rutgers University** Jul 2023 - Sep 2023
Research Assistant New Jersey
Adviser: ***Dong Deng***

- *Find similar data text segment between some documents*
  - **Implement Baseline:** <DataPortraits> **[code]**
  - **Implement Baseline(Python to C++):** <SlimPajama> **[code]**
  - **Assisting In Some Running Experiments:** Such as running slimpajama and get results.
  - **Optimizing Parallelization:** Optimize parallelization for group function.

**4paradigm Company** Jan 2023 - Jul 2023
Research Assistant Singapore
Adviser: ***Mian Lu*** &***Shuhao Zhang***

- **(ML4SYS)***High-Accuracy Low-Latency Stream Window Join with Out-of-Order Data Arrival, Stream Processing System* **[code]**
  - **Implement Baseline:** ICDE 2016<MutipleWay Sliding Window Stream Join> **[code]**
  - **Finding And Organizing Datasets:** using scripts to process the datasets for subsequent experiments

- **Help implement Bayesian variational inference** (transformer + Mathematical method)
- **Optimizing Intra-Window Join:** Optimizing Intra-Window Join using adapative filters.
- **Assisting In Running Experiments Scripts**: Besides, modifying scripts to enhance the visual representation of experimental data.

## WORK EXPERIENCE

**InfiniFlow** Mar 2024 - Apr 2024

Vector Databases Engineer

- Improving the mechanism for recording the oldest visible timestamp to avoid unnecessary access to txn_map.
- Optimizing the cleanup process to allow bulk deletion of files and records, potentially saving significant file IO operations. Additionally, bulk deletion from the buffer manager can reduce lock conflicts.

**Meituan(Top-tier internet company in China)** Apr 2022 - Sep 2022

Software engineer Beijing

- **Meituan's App Short Video project(I built the majority of the foundation for the entire app)**
  - **Data Transimission:** Developed log->kafka->hive to remotely report user video watching behavior to Kafka and synchronize data to Hive tables for recommendation algorithm training.
  - **User Algorithm And Data Refresh:** Improved user experience in poor network environments or for first-time visitors by pulling various data source videos using *Crane* scheduled task to refresh the data periodically.
  - **Implement user group-based resource access:** Through distributed caching *Cellar,* requiring migration of data from Hive data warehouse to a distributed cache Key-Value database due to significant user base growth.

## PROJECTS

**Xv6 operating system kernel**

Xv6 is a unix-like operating system developed by MIT based on Unix v6. After reading the source code and implementation principle of Riscv version XV6, the following functions have been optimized and expanded

**TCP protocol**

The complete TCP protocol is implemented, and the performance is tuned to TCP close to that of Linux kernel, including memory-based ByteStream, stream reassembly. This is also the CS144 experiment of UCB [code]

**Bustub**

Bustub is an open source disk-based relational database of CMU Database Group. It has completed the functions and expansion of the following core components on the basic framework of bustub.[code]

## INTEREST/TALENT

- Machine Learning Systems(Model Inference, Model Compression, Distributed Training)
- Data-Intensive Systems(Data Analytical Systems, Database Systems)

## OTHER

- **Programming language Or Api:** C/C++, CUDA, MPI, OpenMP, Java, Rust, Python
- **Languages**: English(TOEFL 104), Chinese(Native)