

# Hongbin Zhong

☎ +86 18379727287    ✉ [rj986215159@outlook.com](mailto:rj986215159@outlook.com)  
💻 [github:https://github.com/rjzhb](https://github.com/rjzhb)

## RESEARCH INTEREST

---

- Machine Learning For Databases System/Stream Processing System/Data Analytical Workload
- Interactive Data Exploration and Analytics
- New Hardware For Database

## EDUCATION

---

**Northeastern University** 985 Sep 2020 - Jul 2024  
B.S. in Computer Science China  
TOEFL 104

## RESEARCH EXPERIENCE

---

**Columbia University** Jul 2023 - Present  
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 - Oct 2023  
Research Assistant New Jersey  
Adviser: [Dong Deng](#)

- *Find similar data text segment between some documents(Data Cleaning)*
  - **Implement Baseline:** [<DataPortraits> GITHUB](#)
  - **Implement Baseline(Python to C++):** [<SlimPajama> GITHUB](#)
  - **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: [Shuhao Zhang](#)

- *(ML4SYS)High-Accuracy Low-Latency Stream Window Join with Out-of-Order Data Arrival, Stream Processing System* [GITHUB](#)
  - **Implement Baseline:** [ICDE 2016<MutipleWay Sliding Window Stream Join> GITHUB](#)
  - **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 adaptive filters.
- **Assisting In Running Experiments Scripts:** Besides, modifying scripts to enhance the visual representation of experimental data.

## INDUSTRY EXPERIENCE

---

**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.

## 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

[\[code\]](#)

## MANUSCRIPT

---

**[ICDE 2024]** FaDE: Answering “Why?” Made Fast. (Nov 24th)

## OTHER

---

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