

XINYING ZHENG

Email: xz112@illinois.edu | GitHub: <https://github.com/XinyingZheng00> | LinkedIn: [Xinying-Zheng](#)

EDUCATION

University of Illinois Urbana-Champaign

Urbana, USA

PhD in Computer Science

Aug.2023 – Present

- Research Interest: Database and big data systems. Distributed computing. Data processing optimization and advanced data analysis techniques.

Southern University of Science and Technology

Shenzhen, China

Bachelor of Computer Science and Engineering, GPA: 3.93/4.00, Rank: 1/177

Sep.2019 – June 2023

- Summa Cum Laude in Engineering School at SUSTech

RELATED COURSEWORK:

Distributed System, Storage System, Distributed Algorithm, Advanced Data Management, Database System, Operating System, Data Structure and Algorithm Analysis.

TECHNICAL SKILLS

- **Programming Language:** Java, C/C++, Python, SQL, JavaScript
- **Systems, Frameworks, and Tools:** RocksDB, LevelDB, PostgreSQL, AsterixDB, Hadoop, Hive, MapReduce, gRPC, Maven, SpringBoot, Git, GitHub, Ansible, Vue, Jupyter notebook, SSH, Ansible, Kubernetes, Docker

PUBLICATIONS

1. Haotian Liu, Bo Tang, Jiashu Zhang, Yangshen Deng, Xiao Yan, **Xinying Zheng**, [and 13 others]. 2022. GHive: Accelerating Analytical Query Processing in Apache Hive via CPU-GPU Heterogeneous Computing. In *SoCC*. Pages 158-172 [[PDF](#)][[LINK](#)]

WORK EXPERIENCE

- **Graduate Research Assistant at UIUC** Sep.2023 – Present
Supervisor: [Prof. Indranil Gupta](#), [Prof. Yongjoo Park](#)
- **Graduate Teaching Assistant at UIUC** Sep.2024 – Present
Course: CS425 Distributed System
Responsibility: Hold office hours, grade homework, and grade machine problems.
- **Research Intern at UCI.** June 2022 – Sep. 2022
Supervisor: [Prof. Michael J. Carey](#). We revised existing tools to generate UDF-enabled queries, which are used to benchmark the performance of the NoSQL database AsterixDB.
- **Undergraduate Research Assistant at SUSTech** June 2021 – June 2023
Supervisor: [Prof. Bo Tang](#)
- **Undergraduate Teaching Assistant at SUSTech**
Courses: CS207 Digital Logic, CS305 Computer Network, CS302 Operating System

PROJECT EXPERIENCE

Cooperative Compaction for Shared-log-based Distributed System

Oct. 2023 – Present

Role: Project Lead. **Supervisor:** [Prof. Indranil Gupta](#), [Prof. Yongjoo Park](#) at UIUC

- This work allows clients to concurrently compact a shared log-based DB without breaking DB consistency.
- Devised a compaction division technique in the LSM tree storage layout to achieve maximum parallelism.
- Implemented advanced scheduling policies to coordinate compaction and maximize the throughput in the system.

SkyrosFS: An Externally-Synchronous Replicated File System

Oct. 2023 – Dec. 2023

Role: Main developer. **Supervisor:** [Prof. Ramnathan Alagappan](#) at UIUC, [[PDF](#)]

- We designed and implemented SkyrosFS, an externally synchronous replicated file system on top of ext4.

- Utilized the concept of external synchrony in replication, reducing the acknowledgment time for nil-externalizing operations from 2 RTTs to 1 RTT, thereby optimizing system performance.
- Developed an Error Predictor module that anticipates potential errors for nil-external operations, ensuring reliability and stability in the file system operations.
- Achieved 2x speed up for nil-external operations compared with synchronously replicate filesystem atop Filebench[[link](#)]

GHive: Accelerating Apache Hive via CPU-GPU Heterogeneous Computing

June 2021 – June 2022

Role: Main developer. Supervisor: [Prof. Bo Tang](#) at the [Database Group](#) at SUSTech

- This work provides an end-to-end big data query processing system on a CPU-GPU heterogeneous computing environment.
- Implemented GPU-based PTF operator and multiple-keys-join for the Join operator
- Extended STRING and INT data types for the heterogeneous engine
- Scheduled the operators between CPU and GPU based on the execution time
- Tested the system to execute queries in the SSB and TPC-DS under different scale factors
- The research was published as set out above [1]

EPOD: an Edge-resident Framework for Proximity-based Outlier Detection

Sep.2022 – June 2023

Role: Project Lead. Supervisors: [Prof. Bo Tang](#) (at SUSTech) and [Prof. Huan Li](#) (at Aalborg University)

- We proposed a fingerprint-based mechanism to accelerate the outlier detection procedures in the edge computation setting.
- Employed p-stable locality-sensitive hashing (LSH) to generate fingerprints for each edge device, which can avoid concrete pairwise distance computations
- Explored grid index to generate fingerprints and provided a set of pruning strategies to minimize the transmission between edge devices
- The proposed method results in a lower latency and higher energy efficiency.

SIGMOD 2022 Programming Contest, World Finalist (4th out of 55)

Mar.2022 – May 2022

Role: Main developer. Supervisor: [Prof. Bo Tang](#), [Poster](#), [Certificate](#)

The contest aims to develop a blocking system for Entity Resolution on million-level datasets. Our solution consists of two steps: (1) Preprocessing and (2) Blocking.

- In the **preprocessing** step, we used regular rules to extract features from the descriptions, grouped instances with similar features, and embedded entity descriptions using a pre-trained transformer model and indexed the embeddings using HNSW.
- In the **blocking** step, we performed the top-k nearest neighbor search in all groups and output a list of candidate pairs. We then re-ranked the retrieved entity pairs using Euclidean distance and outputted the result until the predetermined output size was reached.