Xiaoke Zhu

Hongshan 6979 Building No.25, Minbao Road, Shenzhen, Guangdong, China

☐ (+86) 15342261274 | ✉ zhuxk@sics.ac.cn | 🌐 https://hsiaoko.github.io | ⌂ hsiaoko

Education

Beihang University   Beijing, China
State Key Laboratory of Software Development   *Sep. 2020 - Sep. 2025*
- Ph.D. in Software Engineering
- Advisor: Prof. Wenfei Fan (Foreign Members of the Chinese Academy of Science, ACM Fellow)

Yunnan University   Kunming, Yunnan
National Pilot School of Software   *Sep. 2017 - Jun. 2020*
- M.Sc. (Engg.) in Software Engineering.
- Advisor: Prof. Shaowen Yao.
- Secondary advisor: Prof. Wei Zhou.

Guilin University of Electronic Technology   Guilin, Guangxi
Computer Engineering Dept.   *Sep. 2013 - Jun. 2016*
- B.E. in Computer Science and Engineering (Self-taught higher education)

Work Experience

Research Fellow, Shenzhen Institute of Computing Sciences   *Sep. 2025 - Present*
- To Add

Research Intern, Shenzhen Institute of Computing Sciences   *Sep. 2021 - Sep. 2025*
- Engaged in research related to databases, data quality, graph computing systems, and GPU-accelerated algorithms. Contributed to several academic papers ($2 \times$ VLDB 2025, VLDB 2023, and Sigmod 2025), $3 \times$ patents and collaborated on algorithm design for database products such as RockDQ and Fishing Fort.

Teaching Fellow, Yunnan University   *Sep. 2018 - Feb. 2019*
- Undergraduate Course: Big data mining and analysis

Research Assistant, Yunnan Radio Monitoring Center & Yunnan University   *Sep. 2017 - May 2020*
- Use deep learning models to classify and detect abnormal signals in radio monitoring. The main tasks include data preprocessing, analysis, simulation, deep learning-based detection, real-time monitoring, and evaluation, ensuring effective identification of abnormal signals.

Papers

$^+$ indicates that author names or major contributors are listed in alphabetical order.
- Wenchao Bai, Wenfei Fan, Shuhao Liu, Kehan Pang, **Xiaoke Zhu**$^+$ and Jiahui Jin. GPU-Accelerated Graph Cleaning with a Single Machine. SIGMOD 2025.
- **Xiaoke Zhu**, Min Xie, Ting Deng and Qi Zhang. HyperBlocker: Accelerating Rule-based Blocking in Entity Resolution using GPUs. PVLDB 2025.
- Yang Liu, Wenfei Fan, Shuhao Liu, **Xiaoke Zhu** and Jianxin Li. A Single Machine System for Querying Big Graphs with PRAM. PVLDB 2025.
- **Xiaoke Zhu**, Qi Zhang, Wei Zhou and Ling Liu. Deep Learning Service for Efficient Data Distribution Aware Sorting. BigData 2024.
- **Xiaoke Zhu**, Yang Liu, Shuhao Liu and Wenfei Fan. MiniGraph: Querying Big Graphs with a Single Machine. PVLDB 2023.
- Ting Deng, Wenfei Fan, Ping Lu, Xiaomeng Luo, **Xiaoke Zhu**$^+$, and Wanhe An. Deep and Collective Entity Resolution in Parallel. ICDE 2022.
- **Xiaoke Zhu**, Qi Zhang, Taining Cheng, Ling Liu, Wei Zhou, and Jing He. DLB: Deep Learning Based Load Balancing, CLOUD 2021.

Research Experience

My research is broadly in the field of graph computing and databases, with an emphasis on optimizing runtime systems for shared-memory and heterogeneous architectures of CPUs/GPUs. This work has been published in SIGMOD, VLDB, ICDE, BigData, CLOUD. A brief summary of my past work can be found below.

Graph Computing Systems
- I have worked on building a high-level programming model and runtime system that can execute applications on shared-memory or out-of-memory architectures with CPUs or GPUs. For out-of-core graph analytics (e.g., PageRank, SSSP), I improved I/O efficiency, and for graph mining (e.g., Graph Data Cleaning, Pattern Matching), I optimized GPU performance. Relevant results were published in [VLDB'23, SIGMOD'25, VLDB'25]

Data Cleaning Systems

- I have worked on improving the performance of data cleaning systems on modern hardware like GPU or on distributed cluster. I have also compared different parallel runtime systems for data cleaning, and identified their performance bottlenecks. Relevant results were published in [ICDE'22, VLDB'25, SIGMOD'25]

AI4DB

- I have leverages machine learning and deep learning model to improve tasks traditionally handled by human database administrators or classical algorithms, enabling more efficient data processing and resource management. Specially I have designed learned models for sorting, load balancing, and scheduling. Relevant results were published in [CLOUD'21, BigData'24]

Awards & Honors

2025   **Shenzhen Distinguished Talent Recruitment(Category C),** 深圳市"引进特聘岗位人才"（C 类）   Shenzhen

Services

Subreviewer   HPCC 2019、ICDE 2024、ICDE 2025、ICDE 2026