Jiacheng Wu

Email: jcwu22@cs.washington.edu
Phone: +1 (206) 258-0585
Homepage: jiacheng-wu.github.io

Computer Science and Engineering,
University of Washington, Box 352350
Seattle, WA 98195-2350

Research Interests

My current research interest lies in database systems and theories, especially big data systems, relational database systems and etc. With my brilliant familiarity of the architectures of several data platforms, especially MySQL, ClickHouse and Apache Spark, I am very interested in building data-intensive systems that can efficiently scale in modern computing environments.

Education

- Ph.D. Computer Science and Engineering, University of Washington, Seattle, WA
  2022.09 - Present. Advisor: Prof. **Dan Suciu**
- M.S. Computer Science and Technology, Tsinghua University, Beijing, China
  2018.09 - 2021.06. GPA: 3.89/4.0; Rank: 11/149
- B.E. Software Engineering, Nankai University, Tianjin, China
  2014.09 - 2018.06. GPA: 93/100; Rank: 1/99

Selected Projects

Topic: The New Architecture of OLTP Databases on Many Cores (Ongoing)
[Work at Institute of Computing Technology, Chinese Academy of Sciences]

- Accelerate real-world RDBMS (e.g., MySQL) by exploiting 1) main memory OLTP engine 2) SQL compilation, and 3) many-core systems.
- Complete a baseline porting main memory OLTP engine (Zen) to work with MySQL and identify bottlenecks of existing databases on many cores.
- Propose and implement a new architecture to deal with bottlenecks, which separates the transaction processing into compilation and execution phases.

Topic: Updatable Learned Indexes [Work at Tsinghua University with Institute of Computing Technology, Chinese Academy of Sciences]

- Propose a brand new updatable learned index with precise key-to-position mapping, eliminating the "last-mile" search in nodes and ensuring lookup/insert cost with theoretical guarantees.
- Devise a linear time algorithm to obtain the model which evenly distributes the mapping as much as possible and a light-weight adjust strategy to keep the index tree height bounded.
- Participate in implementing the index (https://github.com/jiacheng-wu/lipp) for experiments and finish a research paper [3].

Topic: Efficient Parallel Datalog Evaluation [Work at University of California at Los Angeles]

- Propose a new coordination strategy to improve parallel Datalog evaluation in the shared-memory environment by eliminating the requirement of global coordination among all workers.

- Devise and implement a prototype system ($\sim$ 20k of C++ code) with extensive optimizations to improve its overall performance, which outperforms existing solutions by 1 to 2 orders of magnitude.
- Complete a research paper, which is accepted by the SIGMOD 2022 conference [1].

Topic: Large-scale Learning on Spark with Datalog [Work at University of California at Los Angeles]

- Develop a declarative machine learning framework with Datalog as the query interface.
- Propose and implement a series of optimization techniques for aggregates in recursive Datalog programs by modifying the internal query engine of Spark SQL ($\sim$ 1k of Scala/Java code).
- Help in writing a research paper, which is accepted by the VLDB Journal [4].

Topic: Extensible Template Library for Big Data Shuffles. [Work at University of Pennsylvania]

- Provide an extensible shuffle library usable across big data systems and network environments.
- Introduce parameterized shuffle templates populated by sampling functions to dynamically adapt different workloads and layouts.
- Conduct experiments based on Apache Spark by modifying the internal source code ($\sim$ 4k lines of Scala/Java code) and participate in writing a research paper [2].

Topic: Improving Dynamically-Generated Code Performance on Dynamic Binary Translators. [Work at University of Minnesota, Twin Cities]

- Propose an approach to optimize DBT systems for guest applications with DGC which preserves, reuses the DBT code, and retranslates only when needed.
- Implement a prototype based on such an approach on an existing DBT system HQEMU to demonstrate the feasibility ($\sim$ 1k lines of C code) and the effectiveness of the proposed approach.
- Help with the experiment part for a research paper [11].

Publication and Manuscript

1. **Jiacheng Wu**, Jin Wang, Carlo Zaniolo. Optimizing Parallel Recursive Datalog Evaluation on Multicore Machines. Accepted by ACM Special Interest Group in Management Of Data (**SIGMOD**) 2022.
2. **Jiacheng Wu**, Yong Zhang, Shimin Chen, Yu Chen, Jin Wang, Chunxiao Xing. Updatable Learned Index with Precise Positions. Proc. VLDB Endow. 14(8):1276-1288, 2021.
3. Qizhen Zhang, **Jiacheng Wu**, Yucheng Lu, Simran Arora, Ang Chen, Vincent Liu, Boon Thau Loo. Templating Shuffles. Current Status: Manuscript.
4. Jin Wang, **Jiacheng Wu**, Mingda Li, Jiaqi Gu, Ariyam Das, Carlo Zaniolo. Formal Semantics and High Performance in Declarative Machine Learning using Datalog. VLDB J. 30(5):859-881, 2021.
5. Yu Chen, Yong Zhang, **Jiacheng Wu**, Jin Wang, Chunxiao Xing. Revisiting Data Prefetching for Database Systems with Machine Learning Techniques. IEEE International Conference on Data Engineering (**ICDE**) 2021, pages: 2165-2170. (short paper)
6. Yong Zhang, **Jiacheng Wu**, Jin Wang, Chunxiao Xing. A Transformation-Based Framework for KNN Set Similarity Search. IEEE Trans. Knowl. Data Eng. 32(3): 409-423, 2020.
7. **Jiacheng Wu**, Yong Zhang, Yu Chen, Chunxiao Xing. A Progressive Approach for Computing the Earth Mover's Distance. Database Systems for Advanced Applications (**DASFAA**) 2020, pages: 122-138.

8. Jin Wang, Guorui Xiao, Jiaqi Gu, **Jiacheng Wu**, Carlo Zaniolo. RASQL: A Powerful Language and its System for Big Data Applications. ACM International Conference on Management of Data (**SIGMOD**) 2020, pages: 2673-2676. (demo)
9. **Jiacheng Wu**, Yong Zhang, Jin Wang, Chunbin Lin, Yingjia Fu, Chunxiao Xin. Scalable Metric Similarity Join Using MapReduce. IEEE International Conference on Data Engineering (**ICDE**) 2019, pages: 1662-1665. (short paper)
10. Bo Ren, **Jia-Cheng Wu**, Ya-Lei Lv, Ming-Ming Cheng, Shao-Ping Lu. Geometry-Aware ICP for Scene Reconstruction from RGB-D Camera. J. Comput. Sci. Technol. 34(3): 581-593, 2019.
11. Wenwen Wang, **Jiacheng Wu**, Xiaoli Gong, Tao Li, Pen-Chung Yew. Improving Dynamically-Generated Code Performance on Dynamic Binary Translators. 14th ACM SIGPLAN/SIGOPS International Conference on Virtual Execution Environments (**VEE**), pages: 17-30, 2018.
12. Jin Zhang, Yuxuan Li, Chengjun Sun, Haoxiang Yang, **Jiacheng Wu**, Xiaoli Gong. XoT: A Flexible Block I/O Data Transfer Protocol for Xen. Security, Privacy, and Anonymity in Computation, Communication, and Storage (**SpaCCS**) 2017, pages: 791-800. (workshop)

Research Experience

- Research Engineer. Chinese Academy of Sciences. 2021.06 - 2022.06.
  Advisor: Professor Shimin Chen.
- Research Intern. University of California, Los Angeles. 2019.07-2019.09, 2020.01-2020.07 (remotely).
  Advisor: Professor Carlo Zaniolo.
- Research Intern. University of Pennsylvania. 2018.07-2018.09.
  Advisor: Professor Boon Thau Loo and Professor Vincent Liu.
- Research Intern. University of Minnesota, Twin Cities. 2017.07-2017.09.
  Advisor: Professor Pen-Chung Yew.

Honors and Awards

- Siebel Scholar, Class of 2021.
- Outstanding Graduates. Nankai University, 2018.
- National Scholarship for Undergraduate Student (top 1%) three times: 2015, 2016, 2017.

MISC

English Test Score

- TOEFL: 100 (R: 29, L: 27, S: 20, W: 24)
- GRE: 322 (V: 152, Q: 170, AW: 3.5)

Teaching Experience

- TA. Introduction to Database Management System. Tsinghua University, 2019-Autumn.

Academic Service

- PC Member: AAAI 2021

Skills

- Languages: C/C++, Scala, Java, Python, Bash, SQL
- Tools: Spark, MySQL, CUDA, Pytorch, TensorFlow, LaTeX

References

Dan Suciu

Microsoft Endowed Professor.

School of Computer Science and Engineering,

University of Washington,

Email: suciu@cs.washington.edu

Shimin Chen

Professor, Ph.D.

Institute of Computing Technology,

Chinese Academy of Sciences,

Email: chensm@ict.ac.cn

Boon Thau Loo

RCA Professor,

Department of Computer and Information Science,

University of Pennsylvania,

Email: boonloo@seas.upenn.edu

Carlo Zaniolo

N.E. Friedmann Chair Professor, ACM Fellow

Computer Science Department,

University of California, Los Angeles,

Email: zaniolo@cs.ucla.edu

Jin Wang

Researcher, Ph.D.

Megagon Labs,

Mountain View, California, Los Angeles,

Email: jin@ megagon.ai