

Lin Ma

University of Michigan, Ann Arbor
Computer Science and Engineering Division
Bob and Betty Beyster Building
Ann Arbor, MI 48109 USA

Voice: +1-412-519-7097
E-mail: linmacse@umich.edu
Web: <https://web.eecs.umich.edu/~linmacse>
GitHub: <https://github.com/linmagit>

EDUCATION

Ph.D., Computer Science
Carnegie Mellon University
Advisor: [Andy Pavlo](#)

2015 – 2021
Pittsburgh, PA USA

M.Sc., Computer Science
Carnegie Mellon University

2015 – 2018
Pittsburgh, PA USA

B.Sc., Computer Science
Peking University
Advisor: [Bin Cui](#)

2011 – 2015
Beijing, China

PROFESSIONAL EXPERIENCE

Assistant Professor
University of Michigan, Ann Arbor

August 2023 – present
Ann Arbor, MI USA

Software Engineer
Databricks, Inc.

July 2022 – August 2023
Ann Arbor, MI USA

Post Doctoral Fellow
Carnegie Mellon University
Supervisor: [Andy Pavlo](#)

September 2021 – July 2022
Pittsburgh, PA USA

Research Intern
Microsoft Research
Mentors: [Bailu Ding](#), [Sudipto Das](#)

May 2018 – August 2018
Redmond, WA USA

PUBLICATIONS

- [1] Baoqing Cai, Yu Liu, **Lin Ma**, Pingqi Huang, Bingcheng Lian, Ke Zhou, Jia Yuan, Jie Yang, Xiaofan Cai, and Peijun Wu. Scompression: Enhancing database knob tuning efficiency through slice-based oltp workload compression. *Proceedings of the VLDB Endowment*, 2025.
- [2] Yichao Yuan, Advait Iyer, **Lin Ma**, and Nishil Talati. Vortex: Overcoming memory capacity limitations in gpu-accelerated large-scale data analytics. *Proceedings of the VLDB Endowment*, 2025.
- [3] Wan Shen Lim, **Lin Ma**, William Zhang, Matthew Butrovich, Samuel Arch, and Andrew Pavlo. Hit the gym: Accelerating query execution to efficiently bootstrap behavior models for self-driving database management systems. *Proceedings of the VLDB Endowment*, 17(11):3680–3693, 2024.
- [4] Naihao Deng, Zhenjie Sun, Ruiqi He, Aman Sikka, Yulong Chen, **Lin Ma**, Yue Zhang, and Rada Mihalcea. Tables as texts or images: Evaluating the table reasoning ability of llms and mllms. In *Findings of the Association for Computational Linguistics ACL 2024*, pages 407–426, 2024.
- [5] Wan Shen Lim, Matthew Butrovich, William Zhang, Andrew Crotty, **Lin Ma**, Peijing Xu, Johannes Gehrke, and Andrew Pavlo. Database gyms. In *CIDR 2023, Conference on Innovative Data Systems Research*, 2023.

- [6] Matthew Butrovich, Wan Shen Lim, **Lin Ma**, John Rollinson, William Zhang, Yu Xia, and Andrew Pavlo. Tastes great! less filling! high performance and accurate training data collection for self-driving database management systems. In *Proceedings of the 2022 ACM SIGMOD International Conference on Management of Data*, 2022.
- [7] **Lin Ma**, William Zhang, Jie Jiao, Wuwen Wang, Matthew Butrovich, Wan Shen Lim, Prashanth Menon, and Andrew Pavlo. Mb2: Decomposed behavior modeling for self-driving database management systems. In *Proceedings of the 2021 ACM SIGMOD International Conference on Management of Data*, pages 1248–1261, 2021.
- [8] Andrew Pavlo, Matthew Butrovich, **Lin Ma**, Prashanth Menon, Wan Shen Lim, Dana Van Aken, and William Zhang. Make your database system dream of electric sheep: Towards self-driving operation. *Proceedings of the VLDB Endowment*, 14(12):3211–3221, 2021.
- [9] Amadou Ngom, Prashanth Menon, Matthew Butrovich, **Lin Ma**, Wan Shen Lim, Todd C Mowry, and Andrew Pavlo. Filter representation in vectorized query execution. In *Proceedings of the 17th International Workshop on Data Management on New Hardware (DaMoN 2021)*, pages 1–7, 2021.
- [10] Ling Zhang, Matthew Butrovich, Tianyu Li, Andrew Pavlo, Yash Nannapaneni, John Rollinson, Ambarish Zhang, Huachen Balakumar, Daniel Biales, Ziqi Dong, Emmanuel J. Eppinger, Jordi E. Gonzalez, Wan Shen Lim, Jianqiao Liu, **Lin Ma**, Prashanth Menon, Soumil Mukherjee, Tanuj Nayak, Amadou Ngom, Dong Niu, Deepayan Patra, Poojita Raj, Stephanie Wang, Wuwen Wang, Yao Yu, and William Zhang. Everything is a transaction: Unifying logical concurrency control and physical data structure maintenance in database management systems. In *11th Conference on Innovative Data Systems Research, CIDR 2021, Virtual Event, January 11-15, 2021, Online Proceedings*, 2021.
- [11] **Lin Ma**, Bailu Ding, Sudipto Das, and Adith Swaminathan. Active learning for ml enhanced database systems. In *Proceedings of the 2020 ACM SIGMOD International Conference on Management of Data*, pages 175–191, 2020.
- [12] Prashanth Menon, Amadou Ngom, **Lin Ma**, Todd C Mowry, and Andrew Pavlo. Permutable compiled queries: dynamically adapting compiled queries without recompiling. *Proceedings of the VLDB Endowment*, 14(2):101–113, 2020.
- [13] Andrew Pavlo, Matthew Butrovich, Ananya Joshi, **Lin Ma**, Prashanth Menon, Dana Van Aken, Lisa Lee, and Ruslan Salakhutdinov. External vs. internal: an essay on machine learning agents for autonomous database management systems. *IEEE bulletin*, 42(2), 2019.
- [14] **Lin Ma**, Dana Van Aken, Ahmed Hefny, Gustavo Mezerhane, Andrew Pavlo, and Geoffrey J. Gordon. Query-based workload forecasting for self-driving database management systems. In *Proceedings of the 2018 ACM International Conference on Management of Data*, pages 631–645, 2018.
- [15] Andrew Pavlo, Gustavo Angulo, Joy Arulraj, Haibin Lin, Jiexi Lin, **Lin Ma**, Prashanth Menon, Todd C Mowry, Matthew Perron, Ian Quah, et al. Self-driving database management systems. In *CIDR*, 2017.
- [16] **Lin Ma**, Joy Arulraj, Sam Zhao, Andrew Pavlo, Subramanya R Dulloor, Michael J Giardino, Jeff Parkhurst, Jason L Gardner, Kshitij Doshi, and Stanley Zdonik. Larger-than-memory data management on modern storage hardware for in-memory oltp database systems. In *Proceedings of the 12th International Workshop on Data Management on New Hardware*, page 9. ACM, 2016.
- [17] Huachen Zhang, David G Andersen, Andrew Pavlo, Michael Kaminsky, **Lin Ma**, and Rui Shen. Reducing the storage overhead of main-memory oltp databases with hybrid indexes. In *Proceedings of the 2016 International Conference on Management of Data*, pages 1567–1581. ACM, 2016.
- [18] Yingxia Shao, Bin Cui, and **Lin Ma**. Page: a partition aware engine for parallel graph computation. *IEEE Transactions on Knowledge and Data Engineering*, 27(2):518–530, 2015.
- [19] Yingxia Shao, Bin Cui, Lei Chen, **Lin Ma**, Junjie Yao, and Ning Xu. Parallel subgraph listing in a large-scale graph. In *Proceedings of the 2014 ACM SIGMOD International Conference on Management of Data*, pages 625–636. ACM, 2014.

- [20] Yingxia Shao, Junjie Yao, Bin Cui, and **Lin Ma**. Page: A partition aware graph computation engine. In *Proceedings of the 22nd ACM International Conference on Information and Knowledge Management*, pages 823–828. ACM, 2013.

TEACHING

- **Instructor** – CSE 584 Advanced Database Management Systems
University of Michigan, Winter 2026
- **Instructor** – EECS 484 Database Management Systems
University of Michigan, Fall 2025
- **Instructor** – CSE 584 Advanced Database Management Systems
University of Michigan, Winter 2025
- **Instructor** – EECS 484 Database Management Systems
University of Michigan, Fall 2024
- **Instructor** – EECS 484 Database Management Systems
University of Michigan, Winter 2024
- **Instructor** – EECS 584 Advanced Database Management Systems
University of Michigan, Fall 2023
- **Instructor** – 15-445/645 Introduction to Database Systems
Carnegie Mellon University, Fall 2021
- **Head Teaching Assistant** – 15-721 Advanced Database Systems
Carnegie Mellon University, Spring 2019
- **Teaching Assistant** – 15-445/645 Introduction to Database Systems
Carnegie Mellon University, Fall 2018

AWARDS AND SCHOLARSHIPS

- Google ML and Systems Junior Faculty Award – 2025
- The China Computer Federation Outstanding Undergraduate Award – 2015
- China National Scholarship – 2014
- SIGMOD Programming Contest Finalist – 2014
- SIGMOD Travel Award – 2014

SERVICE

To the Profession

- Junior Researcher Symposium co-Chair - SIGMOD 2027
- Panelist - NSF CISE/IIS Division 2025
- Program Committee - SIGMOD 2023, 2024, 2025, 2026, 2027
- Program Committee – VLDB 2022, 2023, 2024, 2025
- Program Committee - CIDR 2024, 2025, 2026
- Web/Information Chair - SIGMOD 2023
- Program Committee – SMDB@ICDE 2022
- Program Committee – AIDB@VLDB 2020, 2021
- External Reviewer – DAPD 2019

- External Reviewer – SIGMOD Demo 2017

To the University

- CSE PhD Admissions Committee – University of Michigan, 2025
- CSE PhD Admissions Committee – University of Michigan, 2024
- CSE PhD Admissions Committee – University of Michigan, 2023
- CSD Faculty Search Committee – Carnegie Mellon University, 2020
- CSD MS Admissions Committee – Carnegie Mellon University, 2018
- Graduate Student Recruitment (Open House) Committee – Carnegie Mellon University, 2018

INVITED TALKS

- **Vortex: Overcoming Memory Capacity Limitations in GPU-Accelerated Large-Scale Data Analytics**
Google, October 10, 2025
Databricks, October 9, 2025
UC Berkeley (Data Systems and Foundations Seminar), October 8, 2025
VLDB, September 3, 2025
Boston University, May 8, 2025
Massachusetts Institute of Technology, May 7, 2025
HEAVY.AI, November 15, 2024
Meta, November 1, 2024
Microsoft Gray Systems Lab, October 22, 2024
Voltron Data, October 15, 2024
- **Self-Driving Databases and the Relevant Research Forefront**
Salesforce (Distinguished AI Speaker Series), April 1, 2024
- **Putting Your Database on Autopilot: Self-Driving Database Management Systems**
Cornell University, September 26, 2023
Google, June 9, 2022
Microsoft Research, March 29, 2022
University of Michigan, Ann Arbor, March 22, 2022
IBM Research, March 10, 2022
University of Maryland, College Park, March 7, 2022
Columbia University, March 2, 2022
Northwestern University, February 23, 2022
Oracle Labs, February 21, 2022
- **NoisePage: The Self-Driving Database Management System**
Ahana, October 19, 2021
University of California, San Diego, October 6, 2021
Facebook, June 4, 2021
Harvard University, April 30, 2021
Columbia University, April 13, 2021
Stanford University (MLSys Seminar), April 8, 2021
Oracle, April 6, 2021
Carnegie Mellon University, March 22, 2021
Centrum Wiskunde & Informatica, March 19, 2021
The University of Chicago, March 17, 2021
University of Washington, March 3, 2021
University of California, Berkeley, February 23, 2021
University of California, Santa Cruz (CSE 215), February 19, 2021
Technical University of Munich, February 18, 2021
Brown University, January 27, 2021

- MB2: Decomposed Behavior Modeling for Self-Driving Database Management Systems
SIGMOD, June 2021
- Active Learning for ML Enhanced Database Systems
SIGMOD, June 2020
- Self-Driving Databases: It All Starts with Workload Forecasting
Percona Live, May 2019
- Efficiently Leveraging B-Instances for Query Plan Predictions
Microsoft Research, August 2018
- Query-based Workload Forecasting for Self-Driving DBMSs
SIGMOD, June 2018
Microsoft Research, May 2018
PDL Retreat, October 2017
- Larger-than-Memory Data Management on Modern Storage Hardware for In-Memory OLTP Database Systems
SIGMOD, June 2016
- The Self-Driving DBMS
PDL Retreat, October 2016
- Multi-Level Anti-Caching for NVM+SSD in H-Store
PDL Retreat, October 2015
- Finalist Presentation of Programming Contest
SIGMOD, June 2014
- Using Less to Do More With Anti-Caching in OLTP Database Systems
Carnegie Mellon University, August 2014