Huanchen Zhang

Tsinghua University
Institute for Interdisciplinary Information Sciences
4-6014 FIT Building, Tsinghua University, Beijing, China 100084

Email: huanchen@tsinghua.edu.cn
Web: http://people.iiis.tsinghua.edu.cn/~huanchen/
Twitter: @huanchenzhang

Education

Carnegie Mellon University — Sept. 2013 - Feb. 2020
Ph.D. in Computer Science Department — *Pittsburgh, PA USA*
Thesis: "Memory-Efficient Search Trees for Database Management Systems"
Advisor: David G. Andersen

University of Wisconsin - Madison — June 2009 - May 2013
B.S. in Computer Engineering, with *Distinctive Scholastic Achievement* — *Madison, WI USA*
2nd and 3rd Major: Computer Sciences and Mathematics
Advisor: Remzi Arpaci-Dusseau

Professional Experience

Assistant Professor — Jan. 2021 - Present
Tsinghua University, Institute for Interdisciplinary Information Sciences (IIIS) — *Beijing, China*

Postdoctoral Research Fellow — April 2020 - Dec. 2020
Snowflake Inc. — *San Mateo, CA USA*

Research Intern — May 2016 - August 2016
Hewlett Packard Labs — *Palo Alto, CA USA*

Awards & Honors

- **2022 ACM China Rising Star Award Honorable Mention**
- 2022 ACM SIGMOD China Rising Star Award
- 2022 CCF Database Committee Distinguished Executive Member
- 2021 World Artificial Intelligence Conference (WAIC) Yunfan Award
- 2021 Excellent Young Scholar (Overseas), National Natural Science Foundation of China
- **2021 ACM SIGMOD Jim Gray Dissertation Award**
- 2020 Communications of the ACM (CACM) Research Highlights
- 2019 ACM SIGMOD Research Highlight Award
- **Best Paper Award, ACM SIGMOD 2018**
- Student Travel Grant, ACM SOSP 2015, USENIX NSDI 2014
- UW-Madison Graduate with Distinctive Scholastic Achievement, 2013
- Edgar H. and Laverne R. Krainer Memorial Scholarship, 2011 & 2012
- UW-Madison Computer Sciences Summer Research Assistant Award, 2012
- The John and Elizabeth Moore Award for Excellence in General Chemistry, 2011

Publications

[1] Yiming Qiao, Peter Boncz, and **Huanchen Zhang**. "Robust Predicate Transfer with Dynamic Execution". In: *Proceedings of the VLDB Endowment* (**VLDB'26**), To Appear.

[2] Yipeng Liu, Renfei Zhou, Jiaqi Yan, and **Huanchen Zhang**. "Workload-Aware Incremental Reclustering in Cloud Data Warehouses". In: *Proceedings of the ACM on Management of Data* (**SIGMOD'26**), To Appear.

[3] Hengrui Zhang, Yulong Hui, Yihao Liu, and **Huanchen Zhang**. "ScaleDoc: Scaling LLM-based Predicates over Large Document Collections". In: *Proceedings of the ACM on Management of Data* (**SIGMOD'26**), To Appear.

[4] Qiuyang Mang, Runyuan He, Suyang Zhong, Xiaoxuan Liu, **Huanchen Zhang**, and Alvin Cheung. "Automated Discovery of Test Oracles for Database Management Systems Using LLMs". In: *Proceedings of the ACM on Management of Data* (**SIGMOD'26**), To Appear.

[5] Youyang Xia, Feng Zhang, Junda Pan, Yihao Liu, Jiawei Guan, **Huanchen Zhang**, and Xiaoyong Du. "L3: A GPU-Native Co-Designed Data Format for Learned Lossless Lightweight Compression" In: *Proceedings of the ACM on Management of Data* (**SIGMOD'26**), To Appear.

[6] Yuxing Han, Yu Lin, Yifeng Dong, Xuanhe Zhou, Xindong Peng, Xinhui Tian, Zhiyuan You, Yingzhong Guo, Xi Chen, Weiping Qu, Tao Meng, Dayue Gao, Haoyu Wang, Liuxi Wei, **Huanchen Zhang**, and Fan Wu. "ByteHouse: ByteDance's Cloud-Native Data Warehouse for Real-Time Multimodal Data Analytics". In: *Proceedings of the ACM on Management of Data* (**SIGMOD'26**), To Appear.

[7] Hengrui Wang, Jiaoyi Zhang, Jiansheng Qiu, Fangzhou Yuan, and **Huanchen Zhang**. "Improving Range Scan Performance in LSM-trees with Group Caching". In: *Proceedings of the ACM on Management of Data* (**SIGMOD'26**), To Appear.

[8] Yuvaraj Chesetti, Navid Eslami, **Huanchen Zhang**, Niv Dayan, and Prashant Pandey. "Aeris Filter: A Strongly and Monotonically Adaptive Range Filter". In: *Proceedings of the ACM on Management of Data* (**SIGMOD'26**), To Appear.

[9] Xinyu Zeng, Ruijun Meng, Martin Prammer, Wes McKinney, Jignesh M. Patel, Andrew Pavlo, and **Huanchen Zhang**. "F3: The Open-Source Data File Format for the Future". In: *Proceedings of the ACM on Management of Data* (**SIGMOD'26**), 3(4): Article 245, 27 pages.

[10] Yulong Hui, Chao Chen, Zhihang Fu, Yihao Liu, Jieping Ye, and **Huanchen Zhang**. "Interact-RAG: Reason and Interact with the Corpus, Beyond Black-Box Retrieval". In: *Proceedings of the 14th International Conference on Learning Representations* (**ICLR'26**), April 2026.

[11] Jiajun Luo, Siyu Lin, Yunpeng Xu, Shengwei Liu, Jin Xia, Dong Liu, Zheng Liu, **Huanchen Zhang**, Teng Ma, and Shuwen Deng. "SHMemora: Protective Key-Value Store on Distributed Shared Memory". In: *Proceedings of the 42nd IEEE International Conference on Data Engineering* (**ICDE'26**), To Appear.

[12] Yu Zhang, Feng Zhang, Yani Liu, **Huanchen Zhang**, Jidong Zhai, Wenchao Zhou, and Xiaoyong Du. "Enabling Tile-Based Direct Query on Adaptively Compressed Data With GPU Acceleration". In: *IEEE Transactions on Parallel and Distributed Systems* (**TPDS'26**), Feb. 2026, vol. 37, pp. 410-426.

[13] Jeronimo Castrillon, Jana Giceva, Yu Hua, Kimberly Keeton, Akhil Shekar, Kevin Skadron, Tianzheng Wang, and **Huanchen Zhang**. "Declarative Memory Services". In: *Proceedings of the 2026 Conference on Innovative Data Systems Research* (**CIDR'26**), January 2026.

[14] Yulong Hui, Yihao Liu, Yao Lu, and **Huanchen Zhang**. "OkraLong: A Flexible Retrieval-Augmented Framework for Long-Text Query Processing". In: *Empirical Methods in Natural Language Processing* (**EMNLP'25**), November 2025.

[15] Yihao Liu, Shaoxuan Tang, Yulong Hui, Hangrui Zhou, and **Huanchen Zhang**. "Selective Late Materialization in Modern Analytical Databases". In: *Proceedings of the VLDB Endowment* (**VLDB'25**), 18.11: 4616-4628.

[16] Weixing Zhou, Yanfeng Zhang, Xinjin Zhou, Zhiyou Wang, Zeshun Peng, Yang Ren, Sihao Li, **Huanchen Zhang**, Guoliang Li, and Ge Yu. "Concurrency Control as a Service". In: *Proceedings of the VLDB Endowment* (**VLDB'25**), 18.9: 2761-2774.

[17] Xiaoyu Fan, Kun Chen, Jiping Yu, Xiaowei Zhu, Yunyi Chen, **Huanchen Zhang**, and Wei Xu. "GORAM: Graph-oriented ORAM for Efficient Ego-centric Queries on Federated Graphs". In: *Proceedings of the VLDB Endowment* (**VLDB'25**), 18.10: 3601-3614.

[18] Xukang Zhang, **Huanchen Zhang**, and Xiaofeng Meng "Accordion: Balancing Performance and Cost in Cloud-Native Data Analysis with Intra-Query Runtime Elasticity". In: *Proceedings of the VLDB Endowment* (**VLDB'25** Demo), 18.12: 5307-5310.

[19] Jiansheng Qiu, Fangzhou Yuan, Mingyu Gao, and **Huanchen Zhang**. "HotRAP: Hot Record Retention and Promotion for LSM-trees with Tiered Storage". In: *Proceedings of the 2025 USENIX Annual Technical Conference* (**ATC'25**). July 2025, pp. 497-511.

[20] Xiang Li, **Huanchen Zhang**, and Mingyu Gao. "TwinStore: Secure Key-Value Stores Made Faster with Hybrid Trusted/Untrusted Storage". In: *Proceedings of the 16th International Symposium on Advanced Parallel Processing Technologies* (**APPT'25**), pp. 321-336.

[21] Yilong Zhao, Mingyu Gao, **Huanchen Zhang**, Fangxin Liu, Gongye Chen, He Xian, Haibing Guan, and Li Jiang. "PUSHtap: PIM-based In-Memory HTAP with Unified Data Storage Format". In: *Proceedings of the 30th ACM International Conference on Architectural Support for Programming Languages and Operating Systems* (**ASPLOS'25**), Volume 3, pp. 179-194.

[22] Junyi Zhao, Kai Su, Yifei Yang, Xiangyao Yu, Paraschos Koutris, and **Huanchen Zhang**. "Debunking the Myth of Join Ordering: Toward Robust SQL Analytics". In: *Proceedings of the ACM on Management of Data* (**SIGMOD'25**), 3(3): Article 146, 28 pages.

[23] Hengrui Wang, Jiansheng Qiu, Fangzhou Yuan, and **Huanchen Zhang**. "Rethinking The Compaction Policies in LSM-trees". In: *Proceedings of the ACM on Management of Data* (**SIGMOD'25**), 3(3): Article 207, 26 pages.

[24] Xukang Zhang, **Huanchen Zhang**, and Xiaofeng Meng. "Intra-Query Runtime Elasticity for Cloud-Native Data Analysis". In: *Proceedings of the ACM on Management of Data* (**SIGMOD'25**), 3(3): Article 178, 28 pages.

[25] Jiaoyi Zhang, Liqiang Peng, Mo Sha, Weiran Liu, Xiang Li, Sheng Wang, Feifei Li, Mingyu Gao, and **Huanchen Zhang**. "Femur: A Flexible Framework for Fast and Secure Querying from Public Key-Value Store". In: *Proceedings of the ACM on Management of Data* (**SIGMOD'25**), 3(3): Article 162, 29 pages.

[26] Yiming Qiao, and **Huanchen Zhang**. "Data Chunk Compaction in Vectorized Execution". In: *Proceedings of the ACM on Management of Data* (**SIGMOD'25**), 3(1): Article 26, 25 pages.

[27] Yiwei Li, Yuxin Jin, Boyu Tian, **Huanchen Zhang**, and Mingyu Gao. "ANSMET: Approximate Nearest Neighbor Search with Near-Memory Processing and Hybrid Early Termination". In: *Proceedings of the 52nd Annual International Symposium on Computer Architecture* (**ISCA'25**). June 2025, pp. 1093-1107

[28] Zhiyu Mei, Wei Fu, Kaiwei Li, Guangju Wang, **Huanchen Zhang**, and Yi Wu. "ReaLHF: Optimized RLHF Training for Large Language Models through Parameter Reallocation". In: *Proceedings of the Eighth Conference on Machine Learning and Systems* (**MLSys'25**), May 2025.

[29] Martin Prammer, Xinyu Zeng, Ruijun Meng, Wes McKinney, **Huanchen Zhang**, Andrew Pavlo, and Jignesh M. Patel. "Towards Functional Decomposition of Storage Formats". In: *Proceedings of the 2025 Conference on Innovative Data Systems Research* (**CIDR'25**), January 2025.

[30] Yulong Hui, Yao Lu, and **Huanchen Zhang**. "UDA: A Benchmark Suite for Retrieval Augmented Generation in Real-world Document Analysis". In: *Proceedings of the Thirty-Eighth Annual Conference on Neural Information Processing Systems* (**NeurIPS'24**), December 2024.

[31] Haowen Dong, Chao Zhang, Guoliang Li, and **Huanchen Zhang** "Cloud-Native Databases: A Survey". In: *IEEE Transactions on Knowledge and Data Engineering* (**TKDE'24**), 36.12: 7772-7791.

[32] Yiming Qiao, Yihan Gao, and **Huanchen Zhang**. "Blitzcrank: Fast Semantic Compression for In-memory Online Transaction Processing". In: *Proceedings of the VLDB Endowment* (**VLDB'24**), 17.10: 2528-2540.

[33] Xinyu Zeng, Yulong Hui, Jiahong Shen, Andrew Pavlo, Wes McKinney, and **Huanchen Zhang**. "An Empirical Evaluation of Columnar Storage Formats". In: *Proceedings of the VLDB Endowment* (**VLDB'24**), 17.2: 148-161.

[34] Xinyu Zeng, Ruijun Meng, Andrew Pavlo, Wes McKinney, and **Huanchen Zhang**. "NULLS!: Revisiting Null Representation in Modern Columnar Formats". In: *Proceedings of the 20th International Workshop on Data Management on New Hardware* (**DaMoN'24**), June 2024, pp. 1-10.

[35] Jiaoyi Zhang, Kai Su, and **Huanchen Zhang**. "Making In-Memory Learned Indexes Efficient on Disk". In: *Proceedings of the ACM on Management of Data* (**SIGMOD'24**), 2.3: Article 151, 26 pages.

[36] Hengrui Wang, Te Guo, Junzhao Yang, and **Huanchen Zhang**. "GRF: A Global Range Filter for LSM-Trees with Shape Encoding". In: *Proceedings of the ACM on Management of Data* (**SIGMOD'24**), 2.3: Article 141, 27 pages.

[37] Shuangyu Cai, Boyu Tian, **Huanchen Zhang**, and Mingyu Gao. "PimPam: Efficient Graph Pattern Matching on Real Processing-in-Memory Hardware". In: *Proceedings of the ACM on Management of Data* (**SIGMOD'24**), 2.3: Article 161, 25 pages.

[38] Yihao Liu, Xinyu Zeng, and **Huanchen Zhang**. "LeCo: Lightweight Compression via Learning Serial Correlations". In: *Proceedings of the ACM on Management of Data* (**SIGMOD'24**), 2.1: Article 65, 28 pages.

[39] Jiake Ge, **Huanchen Zhang**, Boyu Shi, Yuanhui Luo, Yunda Guo, Yunpeng Chai, Yuxing Chen, and Anqun Pan. "SALI: A Scalable Adaptive Learned Index Framework based on Probability Models". In: *Proceedings of the ACM on Management of Data* (**SIGMOD'24**), 1.4: Article 258, 25 pages.

[40] Prashant Pandey, Martín Farach-Colton, Niv Dayan, and **Huanchen Zhang**. "Beyond Bloom: A Tutorial on Future Feature-Rich Filters". In: *Proceedings of the ACM on Management of Data* (**SIGMOD'24**) Tutorial, June 2024.

[41] **Huanchen Zhang**, Yihao Liu, and Jiaqi Yan. "Cost-Intelligent Data Analytics in the Cloud". In: *Proceedings of the 2024 Conference on Innovative Data Systems Research* (**CIDR'24**), January 2024.

[42] Zhiyu Mei, Wei Fu, Guangju Wang, **Huanchen Zhang**, and Yi Wu. "SRL: Scaling Distributed Reinforcement Learning to Over Ten Thousand Cores". In: *Proceedings of the Twelfth International Conference on Learning Representations* (**ICLR'24**), May 2024.

[43] Leon Windheuser, Christoph Anneser, **Huanchen Zhang**, Thomas Neumann, and Alfons Kemper. "AdaCom: Adaptive Compression For Databases". In: *Proceedings of the 27th International Conference on Extending Database Technology* (**EDBT'24**), March 2024.

[44] Jiesong Liu, Feng Zhang, Lv Lu, Chang Qi, Xiaoguang Guo, Dong Deng, Guoliang Li, **Huanchen Zhang**, Jidong Zhai, Hechen Zhang, Yuxing Chen, Anqun Pan, Xiaoyong Du. "G-Learned Index: Enabling Efficient Learned Index on GPU". In: *IEEE Transactions on Parallel and Distributed Systems* (**TPDS'24**), 35.6: 795-812.

[45] Weitao Wan, Feng Zhang Zhang, Chenyang Zhang, Mingde Zhang, Jidong Zhai, Chai, **Huanchen Zhang**, Wei Lu, Yuxing Chen, Haixiang Li, Anqun Pan, and Xiaoyong Du. "Compressed Data Direct Computing for Databases". In: *IEEE Transactions on Knowledge and Data Engineering* (**TKDE'23**), 36.5: 1902-1918.

[46] Junyi Zhao, **Huanchen Zhang**, and Yihan Gao. "Efficient Query Re-optimization with Judicious Subquery Selections". In: *Proceedings of the ACM on Management of Data* (**SIGMOD'23**), 1.2: Article 158, 26 pages.

[47] Ke Wang, Guanqun Yang, Yiwei Li, **Huanchen Zhang**, and Mingyu Gao. "When Tree Meets Hash: Reducing Random Reads for Index Structures on Persistent Memories". In: *Proceedings of the ACM on Management of Data* (**SIGMOD'23**), 1.1: Article 105, 26 pages.

[48] Zheng Chen, Feng Zhang, Jiawei Guan, Jidong Zhai, Xipeng Shen, **Huanchen Zhang**, Wentong Shu, and Xiaoyong Du. "CompressGraph: Efficient Parallel Graph Analytics with Rule-Based Compression". In: *Proceedings of the ACM on Management of Data* (**SIGMOD'23**), 1.1: Article 4, 31 pages.

[49] Hokeun Cha, Xiangpeng Hao, Tianzheng Wang, **Huanchen Zhang**, Aditya Akella, and Xiangyao Yu. "Blink-hash: An Adaptive Hybrid Index for In-Memory Time-Series Databases". In: *Proceedings of the VLDB Endowment* (**VLDB'23**), 16.6: 1235-1248.

[50] Ziwei Wang, Zheng Zhong, Jiarui Guo, Yuhan Wu, Haoyu Li, Tong Yang, Yaofeng Tu, **Huanchen Zhang**, and Bin Cui. "REncoder: A Space-Time Efficient Range Filter with Local Encoder". In: *Proceedings of the 39th IEEE International Conference on Data Engineering* (**ICDE'23**), April 2023, pp. 2036-2049.

[51] Christoph Anneser, Andreas Kipf, **Huanchen Zhang**, Thomas Neumann, and Alfons Kemper. "Adaptive Hybrid Indexes". In: *Proceedings of the 2022 ACM International Conference on Management of Data* (**SIGMOD'22**). June 2022, pp. 1626-1639.

[52] Eric R. Knorr, Baptiste Lemaire, Andrew Lim, Siqiang Luo, **Huanchen Zhang**, Stratos Idreos, Michael Mitzenmacher. "Proteus: A Self-Designing Range Filter". In: *Proceedings of the 2022 ACM International Conference on Management of Data* (**SIGMOD'22**). June 2022, pp. 1670-1684.

[53] Ling Zhang, Matthew Butrovich, Tianyu Li, Yash Nannapanei, Andrew Pavlo, John Rollinson, **Huanchen Zhang**. "Everything is a Transaction: Unifying Logical Concurrency Control and Physical Data Structure Maintenance in Database Management Systems". *Conference on Innovative Data Systems Research* (**CIDR'21**). Jan. 2021.

[54] **Huanchen Zhang**. "Memory-Efficient Search Trees for Database Management Systems". *Ph.D. Thesis.*
ACM SIGMOD Jim Gray Dissertation Award

[55] **Huanchen Zhang**, Hyeontaek Lim, Viktor Leis, David G. Andersen, Michael Kaminsky, Kimberly Keeton, and Andrew Pavlo. "Succinct Range Filters". *Communications of the ACM* (**CACM**). 4 (2021): 166-173.

[56] **Huanchen Zhang**, Lily Liu, David G. Andersen, Michael Kaminsky, Kimberly Keeton, and Andrew Pavlo. "Order-Preserving Key Compression for In-Memory Search Trees". In: *Proceedings of the 2020 ACM International Conference on Management of Data* (**SIGMOD'20**). June 2020, pp. 1601-1615.

[57] **Huanchen Zhang**, Hyeontaek Lim, Viktor Leis, David G. Andersen, Michael Kaminsky, Kimberly Keeton, and Andrew Pavlo. "Succinct Range Filters". *ACM Transactions on Database Systems* (**TODS**). 45.2 (2020): 1-31.

[58] **Huanchen Zhang**, Hyeontaek Lim, Viktor Leis, David G. Andersen, Michael Kaminsky, Kimberly Keeton, and Andrew Pavlo. "Succinct Range Filters". ACM SIGMOD Record, 48.1 (2019): 78-85.

[59] **Huanchen Zhang**, Hyeontaek Lim, Viktor Leis, David G. Andersen, Michael Kaminsky, Kimberly Keeton, and Andrew Pavlo. "SuRF: Practical Range Query Filtering with Fast Succinct Tries". In: *Proceedings of the 2018 ACM International Conference on Management of Data* (**SIGMOD'18**). June 2018, pp. 323–336.
Best Paper Award (1 out of 90 accepted papers)

[60] Ziqi Wang, Andrew Pavlo, Hyeontaek Lim, Viktor Leis, **Huanchen Zhang**, Michael Kaminsky, and David G. Andersen. "Building a Bw-Tree Takes More Than Just Buzz Words". In: *Proceedings of the 2018 ACM International Conference on Management of Data* (**SIGMOD'18**). June 2018, pp. 473–488.

[61] **Huanchen 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* (**SIGMOD'16**). June 2016, pp. 1567–1581.

Non Peer-Reviewed

[62] **Huanchen Zhang**. "The End of the x86 Dominance in Databases?" Abstract. In: *Conference on Innovative Data Systems Research* (**CIDR'19**). Jan. 2019.

Patents

[63] **Huanchen Zhang** and Kimberly Keeton. "Data Storage over Immutable and Mutable Data Stages". Filed Sept. 2017, Granted Sept. 2019.

[64] **Huanchen Zhang** and Kimberly Keeton. "Changing Concurrency Control Modes". Filed May 2017. Patent Pending.

Teaching

- Instructor – Database Systems (Tsinghua 40470414) – Spring 2026, 2025, 2024, 2023, Fall 2021
- Instructor – Introduction to Programming (C/C++) (Tsinghua 30470332) – Fall 2025, 2024, 2023, 2022, Spring 2022
- Instructor – Data Structures in the Real World (Tsinghua 20470102) – Summer 2021
- Head TA & Guest Lecturer – Advanced OS and Distributed Systems (CMU 15-712) – Fall 2017
- Head TA – Database Applications (CMU 15-415/615) – Fall 2016

Students

Current:

- Yihao Liu (Ph.D.)

- Hengrui Wang (Ph.D.)
- Xinyu Zeng (Ph.D.)
- Jiansheng Qiu (Ph.D.)
- Yiming Qiao (Ph.D.)
- Yipeng Liu (Ph.D.)
- Yulong Hui (Ph.D.)
- Kai Su (Ph.D.)
- Jiahong Shen (Ph.D.)
- Ruijun Meng (Ph.D.)
- Xiaoyu Jiang (Ph.D. incoming)
- jiaqi Pan (Ph.D. incoming)
- Weitao Wan (M.S. w/ Mingyu Gao)
- Fangzhou Yuan (M.S.)
- Jiaxin Wang (M.S.)

Alumni:

- Jiaoyi Zhang (Ph.D.)
- Junyi Zhao (Ph.D.)

Service

- SIGMOD Program Committee – 2026, 2023, 2022, 2021, 2020
- VLDB Program Committee – 2027, 2026, 2025, 2024, 2023
- ICDE Program Committee – 2022
- MLSys Program Committee – 2026
- EuroSys Program Committee – 2027
- Jim Gray Award Committee – 2024, 2023, 2022
- FORMATS Co-Chair – 2026
- DaMoN Program Committee – 2026, 2025, 2024
- AIDB Program Committee – 2021
- APSys Program Committee – 2025
- HOPC Program Committee – 2025 2024
- Journal Reviewer – TODS, TKDE, SPE, DISC, KAIS, etc.