Shuhao Zhang

Nationality: Singapore

March 2023

Information Systems Technology and Design (ISTD) Pillar, Singapore University of Technology and Design (SUTD), Singapore.
https://shuhaozhangtony.github.io/
+65 6499 7153
shuhao_zhang@sutd.edu.sg
ShuhaoZhangTony

Education and Qualifications

2014 B. Eng. (Hons) Nanyang Technological University, Singapore
2019 Ph.D. National University of Singapore

Employment

2022– **Technical Consultant**, OpenMLDB team, 4paradigm.
2021– **Assistant Professor**, Information Systems Technology and Design (ISTD), Singapore University of Technology and Design (SUTD).
2020–2021 **Postdoc Research Fellow**, Database Systems and Information Management Group, Technische Universität Berlin (TUB).

Teaching

Courses Taught

Course	Period	Typical class sizes
50.049 Parallel Computing on Multicore Architectures (Sole Teaching)[1]	Spring 2022	20
50.003 Elements of Software Construction (Co-Teaching)	Spring 2021, Spring 2022	200+

Recent Student Teaching Feedback

Sem.	Course	Class Size	#Responses	Subject Rating (/5)	Insturctor Rating (/5)
22S1	50.049	17	17	4.5	4.5
22S1	50.003	207	192	3.4	3.9
21S1	50.003	172	167	2.7[2]	3.2

Research

Ph.D. Thesis

Title: "Scaling Data Stream Processing on Multicore Architectures"
Completed: Oct 2019
Advisor: Dr. Bingsheng He
University: National University of Singapore (NUS)
Impact: review the common bottleneck of modern data stream systems, and develop scale-up solutions.

Research Group and Mission

I am currently leading IntelliStream Group in SUTD. From a high-level point of view, our research goal is to optimize and utilize distributed and parallel stream processing technology to better support existing and emerging big data applications. This is important to improve performance and reduce resource consumption, especially for the network connected world by 5G, IoT, etc. The group has the following research interests:

- Hardware-conscious stream processing algorithms/systems
- Novel stream data management systems/frameworks
- Online machine learning and data stream mining algorithms/systems
- Novel streaming approaches for network function virtualization and software-defined networks

[1] A course developed by me.
[2] Mostly due to course management issues. I was assigned to co-teach this module literally two days after I joined the university.

Research Outputs

- Since 2013 I have authored 20+ papers. A list of selected publications appears on the following page.
- My current research involves large-scale system development and optimization for real-time data stream analytics. Applications include online machine learning, stream data mining, and low-latency query processing.
- I currently (co)supervise 3 PhD students and several visiting students. I have previously (co)supervised another 3 Masters students.
- I have produced 5 open-source systems as a result of my research, all of them are on Github.

Research income

I have acquired (in most cases solely) about S$1.6 million in external research grants since 2021. I am a Principal Investigator for a current Singapore MOE Tier 2 Grant and a Principal Investigator for two current Singapore National Research Foundation Grants. I also have research funding from Temasek Laboratories. Details are listed below. * and # indicate the PI and Co-PI, respectively.

2023-2026	**Shuhao Zhang***, Mian Lu, Tony Quek. "IntelliStream: Towards Highly-Optimized, Ultra-Scalable, Self-adaptive Data Streaming Analytics in the Heterogeneous Multicore IoT Systems". *Funding from Singapore Ministry of Education (MOE) Academic Research Fund (AcRF) Tier 2*	S$650k
2022-2025	**Shuhao Zhang***, Binbin Chen#. "A Stream Processing based NFV Platform for 5G on Modern Multicore Processors". *Funding from National Research Foundation, Singapore and Infocomm Media Development Authority under its Future Communications Research & Development Programme*	S$496k
2022-2025	**Shuhao Zhang***, Mian Lu. "Energy-efficient, Scalable, and Reliable Distributed Green Streaming Machine Learning for Edges". *Funding from National Research Foundation, Singapore and Infocomm Media Development Authority under its Future Communications Research & Development Programme*	S$496k
2022-2025	Meixia Lin*, Das Bikramjit#, Wei Quin Yow#, **Shuhao Zhang#**. "Towards Co-clustering in Big Data: An Optimization Perspective". *Funding from SUTD Kickstarter Initiative (SKI)*	S$477k
2023	**Shuhao Zhang***, Chun Wei Seah, Wei Lu. "Towards Online Continual Pre-Trained Language Model Maintenance". *Funding from TL@SUTD*	S$100k
2022	**Shuhao Zhang***, Wei Lu. "Online Sentiment Learning of Massive Data Streams". *Funding from TL@SUTD*	S$67k
2022-2025	**Shuhao Zhang***. "Revisiting the Algorithms for Clustering Evolving Trajectory Streams". *Funding from SUTD-ZJU (VP)*	S$80k
2021-2024	**Shuhao Zhang***. "Efficient Intra-Window Join on the Multicore IoT systems". *Funding from START-UP RESEARCH GRANT (SRG)*	S$100k

Publications

Author notations: * denotes the author is a student advised by me. # denotes the author is a staff advised by me.

Five Selected Articles

1. Mao#, Y, J Zhao, H Liu, **Shuhao Zhang**, and V Markl (2023). MorphStream: Adaptive Scheduling for Scalable Transactional Stream Processing on Multicores. In: *Proceedings of the 2023 International Conference on Management of Data (SIGMOD)*. SIGMOD '23. Seattle, WA, USA: Association for Computing Machinery.
2. **Shuhao Zhang**, Y Mao, J He, PM Grulich, S Zeuch, B He, RTB Ma, and V Markl (2021). Parallelizing Intra-Window Join on Multicores: An Experimental Study. In: *Proceedings of the 2021 International Conference on Management of Data (SIGMOD)*. SIGMOD '21. Xi'an, Shaanxi, China: Association for Computing Machinery.
3. **Shuhao Zhang**, Y Wu, F Zhang, and B He (2020). Towards Concurrent Stateful Stream Processing on Multicore Processors. In: *2020 IEEE 36th International Conference on Data Engineering (ICDE)*, pp.1537–1548.
4. **Shuhao Zhang**, J He, AC Zhou, and B He (2019). BriskStream: Scaling Data Stream Processing on Shared-Memory Multicore Architectures. In: *Proceedings of the 2019 International Conference on Management of Data (SIGMOD)*. SIGMOD '19. Amsterdam, Netherlands: Association for Computing Machinery, pp.705–722. `https://doi.org/10.1145/3299869.3300067`.
5. **Shuhao Zhang**, B He, D Dahlmeier, AC Zhou, and T Heinze (2017). Revisiting the Design of Data Stream Processing Systems on Multi-Core Processors. In: *2017 IEEE 33rd International Conference on Data Engineering (ICDE)*, pp.659–670.

Refereed Research Papers

1. Mao#, Y, J Zhao, H Liu, **Shuhao Zhang**, and V Markl (2023). MorphStream: Adaptive Scheduling for Scalable Transactional Stream Processing on Multicores. In: *Proceedings of the 2023 International Conference on Management of Data (SIGMOD)*. SIGMOD '23. Seattle, WA, USA: Association for Computing Machinery.

2. Wang*, X, Z Wang*, Z Wu#, S Zhang, X Shi, and L Lu (2023). Data Stream Clustering: An In-depth Empirical Study. In: *Proceedings of the 2023 International Conference on Management of Data (SIGMOD)*. SIGMOD '23. Seattle, WA, USA: Association for Computing Machinery.
3. Zeng*, X and **Shuhao Zhang** (2023). A Hardware-Conscious Stateful Stream Compression Framework for IoT Applications (Vision). In: *Proceedings of the 2023 International Conference on Distributed and Event-Based Systems (DEBS)*.
4. Zeng*, X and **Shuhao Zhang** (2023). Parallelizing Stream Compression for IoT Applications on Asymmetric Multicores. In: *Proceedings of the 2023 IEEE 39rd International Conference on Data Engineering (ICDE)*.
5. Zhang, H, X Zeng*, **Shuhao Zhang**, X Liu, M Lu, and Z Zheng (2023). Scalable Online Interval Join on Modern Multicore Processors in OpenMLDB. In: *2023 IEEE 39rd International Conference on Data Engineering (ICDE)*.
6. Zhang, Y, F Zhang, H Li, **Shuhao Zhang**, and X Du (2023). CompressStreamDB: Fine-Grained Adaptive Stream Processing without Decompression. In: *Proceedings of the 2023 IEEE 39rd International Conference on Data Engineering (ICDE)*.
7. Meftah, S, **Shuhao Zhang**, B Veeravalli, and KMM Aung (2022). Revisiting the Design of Parallel Stream Joins on Trusted Execution Environments. *Algorithms* **15**(6).
8. Xu, Q, F Zhang, M Zhang, J Zhai, B He, C Yang, **Shuhao Zhang**, J Lin, H Liu, and X Du (2022). Payment behavior prediction on shared parking lots with TR-GCN. *The VLDB Journal*.
9. **Shuhao Zhang**, Y Mao, J He, PM Grulich, S Zeuch, B He, RTB Ma, and V Markl (2021). Parallelizing Intra-Window Join on Multicores: An Experimental Study. In: *Proceedings of the 2021 International Conference on Management of Data (SIGMOD)*. SIGMOD '21. Xi'an, Shaanxi, China: Association for Computing Machinery.
10. Zhang, F, Y Liu, N Feng, C Yang, J Zhai, **Shuhao Zhang**, B He, J Lin, X Zhang, and X Du (2021). Periodic Weather-Aware LSTM with Event Mechanism for Parking Behavior Prediction. *IEEE Transactions on Knowledge and Data Engineering*, 1–1.
11. Zhang, F, C Zhang, L Yang, C Yang, **Shuhao Zhang**, B He, W Lu, and X Du (2021). Fine-Grained Multi-Query Stream Processing on Integrated Architectures. *IEEE Transactions on Parallel and Distributed Systems (TPDS)* **32**(9), 2303–2320.
12. **Shuhao Zhang**, Y Wu, F Zhang, and B He (2020). Towards Concurrent Stateful Stream Processing on Multicore Processors. In: *2020 IEEE 36th International Conference on Data Engineering (ICDE)*, pp.1537–1548.
13. **Shuhao Zhang**, F Zhang, Y Wu, B He, and P Johns (2020). Hardware-Conscious Stream Processing: A Survey. *SIGMOD Rec.* **48**(4), 18–29.
14. Zeuch, S, ET Zacharatou, **Shuhao Zhang**, X Chatziliadis, A Chaudhary, BD Monte, D Giouroukis, PM Grulich, A Ziehn, and V Markl (2020). NebulaStream: Complex Analytics Beyond the Cloud. *Open Journal of Internet Of Things (OJIOT)* **6**(1), 66–81.
15. Zhang, F, N Feng, Y Liu, C Yang, J Zhai, **Shuhao Zhang**, B He, J Lin, and X Du (2020). PewLSTM: Periodic LSTM with Weather-Aware Gating Mechanism for Parking Behavior Prediction. In: *International Joint Conference on Artificial Intelligence(IJCAI)*.
16. Zhang, F, L Yang, **Shuhao Zhang**, B He, W Lu, and X Du (2020). FineStream: Fine-Grained Window-Based Stream Processing on CPU-GPU Integrated Architectures. In: *2020 USENIX Annual Technical Conference (USENIX ATC 20)*. USENIX Association, pp.633–647. `https://www.usenix.org/conference/atc20/presentation/zhang-feng`.
17. Ang*, J, T Fu*, J Paul, **Shuhao Zhang**, B He, TSD Wenceslao, and SY Tan (2019). TraV: An Interactive Exploration System for Massive Trajectory Data. In: *2019 IEEE Fifth International Conference on Multimedia Big Data (BigMM)*, pp.309–313.
18. **Shuhao Zhang**, J He, AC Zhou, and B He (2019). BriskStream: Scaling Data Stream Processing on Shared-Memory Multicore Architectures. In: *Proceedings of the 2019 International Conference on Management of Data (SIGMOD)*. SIGMOD '19. Amsterdam, Netherlands: Association for Computing Machinery, pp.705–722. `https://doi.org/10.1145/3299869.3300067`.
19. **Shuhao Zhang**, B He, D Dahlmeier, AC Zhou, and T Heinze (2017). Revisiting the Design of Data Stream Processing Systems on Multi-Core Processors. In: *2017 IEEE 33rd International Conference on Data Engineering (ICDE)*, pp.659–670.
20. **Shuhao Zhang**, HT Vo, D Dahlmeier, and B He (2017). Multi-Query Optimization for Complex Event Processing in SAP ESP. In: *2017 IEEE 33rd International Conference on Data Engineering (ICDE)*, pp.1213–1224.
21. Zhang, F, J Zhai, B He, **Shuhao Zhang**, and W Chen (2017). Understanding Co-Running Behaviors on Integrated CPU/GPU Architectures. *IEEE Transactions on Parallel and Distributed Systems* **28**(3), 905–918.
22. Tang, S, B He, **Shuhao Zhang**, and Z Niu (2016). Elastic Multi-resource Fairness: Balancing Fairness and Efficiency in Coupled CPU-GPU Architectures. In: *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*, pp.875–886.

23. Wang, Z, **Shuhao Zhang**, B He, and W Zhang (2016). Melia: A MapReduce Framework on OpenCL-Based FPGAs. *IEEE Transactions on Parallel and Distributed Systems* **27**(12), 3547–3560.
24. Zhang, F, J Zhai, W Chen, B He, and **Shuhao Zhang** (2015). To Co-run, or Not to Co-run: A Performance Study on Integrated Architectures. In: *2015 IEEE 23rd International Symposium on Modeling, Analysis, and Simulation of Computer and Telecommunication Systems*, pp.89–92.
25. He, J, **Shuhao Zhang**, and B He (2014). In-Cache Query Co-Processing on Coupled CPU-GPU Architectures. *Proceedings of the VLDB Endowment* **8**(4), 329–340.
26. **Shuhao Zhang**, J He, B He, and M Lu (2013). OmniDB: Towards Portable and Efficient Query Processing on Parallel CPU/GPU Architectures. *Proceedings of the VLDB Endowment* **6**(12), 1374–1377.