Zhang Shuhao

🏠 https://shuhaozhangtony.github.io/
✉ shuhao_zhang@hust.edu.cn
⌂ https://github.com/intellistream

Employment

2023–2024	**Assistant Professor**, College of Computing and Data Science, Nanyang Technological University (NTU).
2021–2023	**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).

Consultancy Position

2022–2023	**Technical Consultant** (Part-Time), OpenMLDB team, 4paradigm.

Education

2019	PhD.	Computer Science (Database Systems)	National University of Singapore
2014	B.Eng. (1st Hons)	Computer Engineering	Nanyang Technological University, Singapore

Awards/Patents

2018-2019	Research Achievement Award, School of Computing, National University of Singapore
2017-2018	Research Achievement Award, School of Computing, National University of Singapore
2019/12/31	Efficient execution of data stream processing systems on multi-core processors, US Patent (10521432)
2018/4/24	Multi-query optimizer for complex event processing, US Patent (9953056)

Service

I have helped in organizing various events.

Years	Event Name	Role
2023	National Research Foundation - Foundational Research Capabilities Study (*NRF-FRC, Singapore*)	Area Leader
2023	National Supercomputing Centre HPC AI Innovation Challenge 2023 (*NSCC, Singapore*)	Organising Committee
2022	Singapore Data Science Consortium (*SDSC, Singapore*) PhD Dissertation Fellowship Evaluation	Evaluator

I have regularly served on the program committee of various international conferences:

Years	Conference Name	Role
2022-2025	*DASFAA*: Database Systems for Advanced Applications	Program Committee Member
2024	*ICDE*: IEEE International Conference on Data Engineering	Chair of TKDE Posters Track & PC Member
2023-2024	*ADMA*: International Conference on Advanced Data Mining and Applications	Program Committee Member
2024	*SDM*: SIAM International Conference on Data Mining	Program Committee Member
2023	NRF FRC Green Computing - Green Edge Workshop	Co-chair
2021-2023	*SC*: International Conference for High Performance Computing, Networking, Storage, and Analysis	Program Committee Member
2021-2023	*ICPP*: International Conference on Parallel Processing	Program Committee Member
2022-2023	*KDD*: ACM SIGKDD Conference on Knowledge Discovery and Data Mining	Program Committee Member
2023	*EDBT*: International Conference on Extending Database Technology	Program Committee Member
2020-2021, 2023	*HiPC*: IEEE International Conference on High Performance Computing, Data, and Analytics	Program Committee Member
2022	*DEBS*: International Conference on Distributed and Event-based Systems	Program Committee Member
2020-2021	*ICDCS*: International Conference on Distributed Computing Systems	Program Committee Member
2021	*Apsys*: ACM Asia-Pacific Workshop on Systems	Program Committee Member
2021	*CCGrid*: International Symposium on Cluster, Cloud and Internet Computing	Program Committee Member
2019	*CIKM*: ACM International Conference on Information and Knowledge Management	Program Committee Member
2019	*CCIS*: IEEE International Conference on Cloud Computing and Intelligence Systems	Program Committee Member

I regularly served as a reviewer for various peer-reviewed journals:

➤ IEEE Transactions on Knowledge and Data Engineering (TKDE)

➤ IEEE Transactions on Cloud Computing (TCC)

➤ ACM Transactions on Database Systems (TODS)

➤ ACM Transactions on Parallel Computing

➤ IEEE Transactions on Parallel and Distributed Systems (TPDS)

➤ IEEE Transactions on Services Computing (TSC)

➤ Information Systems (IS)

➤ Pervasive and Mobile Computing (PMC)

➤ Journal of Computer Science and Technology (JCST)

➤ Future Generation Computer Systems (FGCS)

➤ Computers & Security (CS)

I have also served as a proposal evaluator for research grants:

➤ Future Communications Programme, Singapore (Proposal of 3 Years)

➤ Israel Science Foundation Personal Research Grants, Israel (Proposal of 4 Years)

Publications

Citations: 948, h-index: 16.

Five Selected Articles

1. Zeng, X, Zhang, Shuhao*, H Zhong, H Zhang, M Lu, Z Zheng, and Y Chen (2024). PECJ: Stream Window Join on Disorder Data Streams with Proactive Error Compensation. *Proc. ACM Manag. Data* **2**(1), V2mod013:1–V2mod013:24. DOI: `10.1145/3639268`.

2. Mao, Y, J Zhao, H Liu, Zhang, Shuhao*, and V Markl (2023). MorphStream: Adaptive Scheduling for Scalable Transactional Stream Processing on Multicores. *Proc. ACM Manag. Data* **1**(1). 26 pages. DOI: `10.1145/3588913`.

3. Zhang, Shuhao*, 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. 13 pages. Xi'an, Shaanxi, China: Association for Computing Machinery.

4. Zhang, Shuhao*, 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. 18 pages. Amsterdam, Netherlands: Association for Computing Machinery, pp.705–722. DOI: `10.1145/3299869.3300067`.

5. Zhang, Shuhao*, 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)*. 12 pages, pp.659–670. DOI: `10.1109/ICDE.2017.119`.

Refereed Research Papers

1. Li, J, H Yuan, G Cong, HM Kiah, and Zhang, Shuhao (2025). MAST: Towards Efficient Analytical Query Processing on Point Cloud Data. *International Conference on Management of Data (SIGMOD'25)*, *12 pages*.

2. Tang, X, F Zhang, Zhang, Shuhao, Y Liu, J Cheng, B He, and X Du (2025). Bi-Probe: Adaptive Sampling-Aware Intra-Window Join. *International Conference on Management of Data (SIGMOD'25)*, *12 pages*.

3. Li, Y, D Song, C Zhou, Y Tian, H Wang, Z Yang, and Zhang, Shuhao* (2024). A Framework of Knowledge Graph-Enhanced Large Language Model Based on Question Decomposition and Atomic Retrieval. In: *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*. 13 pages, long paper, Findings.

4. Wang, R, C Xu, Zhang, Shuhao, F Ye, Y Tang, S Tang, H Zhang, W Du, and X Zhang (2024). MatSwarm: trusted swarm transfer learning driven materials computation for secure big data sharing. *Nature Communications* **15**(1), 1–14.

5. Wang, X, Z Wang, and Zhang, Shuhao* (2024). MOStream: A Modular and Self-Optimizing Data Stream Clustering Algorithm. *IEEE International Conference on Data Mining*, *10 pages*.

6. Xiao, Y, S Chen, AC Zhou, Zhang, Shuhao, Y Wang, R Mao, and X Yang (2024). Low-Latency Video Conferencing via Optimized Packet Routing and Reordering. In: *32nd IEEE/ACM International Symposium on Quality of Service, IWQoS 2024, Guangzhou, China, June 19-21, 2024*. IEEE, pp.1–10. DOI: `10.1109/IWQOS61813.2024.10682858`.

7. Xu, Y, D Zhang, Zhang, Shuhao, S Wu, Z Feng, and G Chen (2024). Predictive and Near-Optimal Sampling for View Materialization in Video Databases. *Proceedings of the ACM on Management of Data* **2**(1), 1–27.

8. Yang, D, Z Xiao, D Zhang, Zhang, Shuhao, J Cao, and G Chen (2024). PREACT: Predictive Resource Allocation for Bursty Workloads in a Co-located Data Center. In: *Proceedings of the 53rd International Conference on Parallel Processing, ICPP 2024, Gotland, Sweden, August 12-15, 2024*. ACM, pp.722–731. DOI: `10.1145/3673038.3673135`.

9. Zeng, X, W Jiang, and Zhang, Shuhao* (2024). LibAMM: Empirical Insights into Approximate Computing for Accelerating Matrix Multiplication. In: *Advances in Neural Information Processing Systems*. Curran Associates, Inc.

10. Zeng, X and Zhang, Shuhao* (2024). CStream: Parallel Data Stream Compression on Multicore Edge Devices. *IEEE Transactions on Knowledge and Data Engineering (TKDE), 14 pages*. Available at `https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=10506068&tag=1`.

11. Zeng, X, Zhang, Shuhao*, H Zhong, H Zhang, M Lu, Z Zheng, and Y Chen (2024). PECJ: Stream Window Join on Disorder Data Streams with Proactive Error Compensation. *Proc. ACM Manag. Data* **2**(1), V2mod013:1–V2mod013:24. DOI: `10.1145/3639268`.

12. Zhang, Y, F Zhang, H Li, Zhang, Shuhao, X Guo, Y Chen, A Pan, and X Du (2024). Data-Aware Adaptive Compression for Stream Processing. *IEEE Transactions on Knowledge and Data Engineering*.

13. Zhang, Shuhao, J Soto, and V Markl (2024). A survey on transactional stream processing. *VLDB J.* **33**(2), 451–479. DOI: `10.1007/S00778-023-00814-Z`.

14. Zhao, J, H Liu, Zhang, Shuhao, Z Duan, X Liao, H Jin, and Y Zhang (2024). Fast Parallel Recovery for Transactional Stream Processing on Multicores. In: *40th IEEE International Conference on Data Engineering, ICDE 2024, Utrecht, The Netherlands, May 13-16, 2024*. IEEE, pp.1478–1491. DOI: `10.1109/ICDE60146.2024.00122`.

15. Mao, Y, J Zhao, H Liu, Zhang, Shuhao*, and V Markl (2023). MorphStream: Adaptive Scheduling for Scalable Transactional Stream Processing on Multicores. *Proc. ACM Manag. Data* **1**(1). 26 pages. DOI: `10.1145/3588913`.

16. Wang, X, Z Wang, Z Wu, Zhang, Shuhao*, X Shi, and L Lu (2023). Data Stream Clustering: An In-Depth Empirical Study. *Proc. ACM Manag. Data* **1**(2). 26 pages. DOI: `10.1145/3589307`.

17. Wu, Y, KT Sharma, CW Seah, and Zhang, Shuhao* (2023). A Co-Training Framework for Adaptive Online Sentiment Analysis in Dynamic Data Streams. In: *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*. 13 pages, main track.

18. Zeng, X and Zhang, Shuhao* (2023). A Hardware-Conscious Stateful Stream Compression Framework for IoT Applications (Vision). In: *Proceedings of the 17th ACM International Conference on Distributed and Event-Based Systems*. DEBS '23. 6 pages. Neuchatel, Switzerland: Association for Computing Machinery, pp.7–12. DOI: `10.1145/3583678.3596885`.

19. Zeng, X and Zhang, Shuhao* (2023). Parallelizing Stream Compression for IoT Applications on Asymmetric Multicores. In: *Proceedings of the 2023 IEEE 39rd International Conference on Data Engineering (ICDE)*. 12 pages.

20. Zhang, H, X Zeng, Zhang, Shuhao, 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)*. 12 pages, industry track.

21. Zhang, Y, F Zhang, H Li, Zhang, Shuhao, 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)*. 12 pages.

22. Meftah, S, Zhang, Shuhao, B Veeravalli, and KMM Aung (2022). Revisiting the Design of Parallel Stream Joins on Trusted Execution Environments. *Algorithms* **15**(6). 20 pages. DOI: `10.3390/a15060183`.

23. Xu, Q, F Zhang, M Zhang, J Zhai, B He, C Yang, Zhang, Shuhao, J Lin, H Liu, and X Du (2022). Payment behavior prediction on shared parking lots with TR-GCN. *The VLDB Journal*. 24 pages. DOI: `10.1007/s00778-021-00722-0`.

24. Zhang, F, Y Liu, N Feng, C Yang, J Zhai, Zhang, Shuhao, 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*. 14 pages, 1–1. DOI: `10.1109/TKDE.2021.3070202`.

25. Zhang, F, C Zhang, L Yang, C Yang, Zhang, Shuhao, 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). 14 pages, 2303–2320. DOI: `10.1109/TPDS.2021.3066407`.

26. Zhang, Shuhao*, 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. 13 pages. Xi'an, Shaanxi, China: Association for Computing Machinery.

27. Zeuch, S, ET Zacharatou, Zhang, Shuhao, 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). 16 pages, 66–81.

28. Zhang, F, N Feng, Y Liu, C Yang, J Zhai, Zhang, Shuhao, 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)*. 7 pages.

29. Zhang, F, L Yang, Zhang, Shuhao, 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)*. 12 pages. USENIX Association, pp.633–647.

30. Zhang, Shuhao*, 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)*. 12 pages, pp.1537–1548. DOI: `10.1109/ICDE48307.2020.00136`.

31. Zhang, Shuhao*, F Zhang, Y Wu, B He, and P Johns (2020). Hardware-Conscious Stream Processing: A Survey. *SIGMOD Rec.* **48**(4). 12 pages, 18–29. DOI: `10.1145/3385658.3385662`.

32. Ang, J, T Fu, J Paul, Zhang, Shuhao*, 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)*. 4 pages, pp.309–313. DOI: `10.1109/BigMM.2019.000-4`.

33. Zhang, Shuhao*, 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. 18 pages. Amsterdam, Netherlands: Association for Computing Machinery, pp.705–722. DOI: `10.1145/3299869.3300067`.

34. Zhang, F, J Zhai, B He, Zhang, Shuhao, and W Chen (2017). Understanding Co-Running Behaviors on Integrated CPU/GPU Architectures. *IEEE Transactions on Parallel and Distributed Systems* **28**(3). 14 pages, 905–918. DOI: `10.1109/TPDS.2016.2586074`.

35. Zhang, Shuhao*, 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)*. 12 pages, pp.659–670. DOI: `10.1109/ICDE.2017.119`.

36. Zhang, Shuhao*, 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)*. 12 pages, pp.1213–1224. DOI: `10.1109/ICDE.2017.166`.

37. Tang, S, B He, Zhang, Shuhao, 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*. 12 pages, pp.875–886. DOI: `10.1109/SC.2016.74`.

38. Wang, Z, Zhang, Shuhao, B He, and W Zhang (2016). Melia: A MapReduce Framework on OpenCL-Based FPGAs. *IEEE Transactions on Parallel and Distributed Systems* **27**(12). 14 pages, 3547–3560. DOI: `10.1109/TPDS.2016.2537805`.

39. Zhang, F, J Zhai, W Chen, B He, and Zhang, Shuhao (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*. 4 pages, pp.89–92. DOI: `10.1109/MASCOTS.2015.27`.

40. He, J, Zhang, Shuhao, and B He (2014). In-Cache Query Co-Processing on Coupled CPU-GPU Architectures. *Proceedings of the VLDB Endowment* **8**(4). 12 pages, 329–340. DOI: `10.14778/2735496.2735497`.

41. Zhang, Shuhao, 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). 4 pages, 1374–1377. DOI: `10.14778/2536274.2536319`.

Book chapters

1. Zhang, Shuhao, M Plauth, F Eberhardt, A Polze, J Lehmann, G Sejdiu, H Jabeen, L Servadei, C Möstl, F Bär, et al. (2020). *HPI Future SOC Lab–Proceedings 2017*. Universitätsverlag Potsdam.

Appendix

Research income

I have acquired (in most cases solely) about 14 million RMB in external research grants since 2021. I am a Principal Investigator for Outstanding Young Talents Programme, China. I was a Principal Investigator for a 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.

2024-2026 (On-Going)	**Shuhao Zhang***. "Principal Investigator for Outstanding Young Talents Programme, China". *National Natural Science Foundation of China*	~RMB$6M
2023-2026 (Terminated)	**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$500k
2022-2025 (Transferred)	**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 (Transferred)	**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 (Quit)	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
2023-2023 (Transferred)	**Shuhao Zhang***, Chun Wei Seah, Wei Lu. "Towards Online Continual Pre-Trained Language Model Maintenance". *Funding from TL@SUTD*	S$100k
2022-2022 (Completed)	**Shuhao Zhang***, Wei Lu. "Online Sentiment Learning of Massive Data Streams". *Funding from TL@SUTD*	S$67k
2022-2025 (Terminated)	**Shuhao Zhang***. "Revisiting the Algorithms for Clustering Evolving Trajectory Streams". *Funding from SUTD-ZJU (VP)*	S$80k
2021-2024 (Terminated)	**Shuhao Zhang***. "Efficient Intra-Window Join on the Multicore IoT systems". *Funding from START-UP RESEARCH GRANT (SRG)*	S$100k

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, Summer 2023	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
23S2	50.003	212	199	3.8	4.1
22S1	50.003	207	192	3.4	3.9
21S1	50.003	172	167	2.7[2]	3.2

---

[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.

Referees

The following referees will email the recommendation letters directly if asked:

- ➤ **Prof. Bingsheng He** (PhD Supervisor) - National University of Singapore
  Email: hebs@comp.nus.edu.sg

- ➤ **Prof. Volker Markl** (Postdoc Advisor) - Technische Universität Berlin
  Email: volker.markl@tu-berlin.de

- ➤ **Prof. Xiaoyong Du** (Senior Collaborator) - Renmin University of China
  Email: duyong@ruc.edu.cn