Curriculum Vitae

HAI LAN Male, Born in Sep 1993

Tel: 0410270039 E-Mail: hai.lan@student.rmit.edu.au

EDUCATION

May. 2021 – Present	PhD of Computer Science RMIT University, Melbourne, Australia Supervisor: Prof. Zhifeng Bao, Prof. J. Shane Culpepper
Sep. 2017 – May. 2020	Master of Computer Science and Technology Wuhan University, Wuhan, China（TOP-10 in China） Rank in Computer Science: USNews: 49 Supervisor: Dr. Yuwei Peng Avg Score: 89.3/100
Sep. 2013 – June. 2017	Bachelor of Computer Science and Technology Wuhan University, Wuhan, China Avg Score: 89.6/100 Rank 12/220 First-Class Honors

RESEARCH INTERESTS

database system, query optimization, index tuning, learned index, spatio-temporal data processing

WORKING EXPERIENCE

- ✧ Research Intern Alibaba Damo Academy (Aug. 2020 – Feb. 2021)
- ✧ Software Engineer at PingCAP (Nov. 2018 – Apr. 2019)
- ✧ Teaching Assistant in Computer School of Wuhan University
  - Database System Implementation Seg. 2017- Jan. 2018
  - Database System Concept Seg. 2018- Jan. 2019

PUBLICATIONS

1. **Hai Lan**, Jiong Xie, Zhifeng Bao, Feifei Li, Wei Tian, Fang Wang, Sheng Wang, Ailin Zhang. VRE: A Versatile, Robust, and Economical Trajectory Data System. VLDB 2022. (CORE A*[1])
2. Juntao Zhang, **Hai Lan**, Xiandi Yang, Shuaichao Zhang, Wei Song, Zhiyong Peng. Weakly supervised setting for learning concept prerequisite relations using multi-head attention variational graph auto-encoders. Knowledge-Based Systems. Volume 247, 2022 (CORE B)
3. Yuwei Peng, **Hai Lan**, Mingling Yue, Yu Xue. Multipurpose watermarking for vector map protection and authentication. Multimedia Tools Appl. 77(6): 7239-7259(2018) (CORE B)
4. **Hai Lan**, Yuwei Peng. Reversible Fragile Watermarking for Fine-Grained Tamper Localization in Spatial Data. Australasian Database Conference (ADC) 2017: 233 – 247 (Australasian B)
5. **Hai Lan**, Zhifeng Bao, Yuwei Peng: An Index Advisor Using Deep Reinforcement Learning.

[1] The Computing Research and Education Association of Australasia (CORE), http://www.core.edu.au/

CIKM 2020: 2105-2108 (CORE A)

6. **Hai Lan**, Zhifeng Bao, Yuwei Peng: A Survey on Advancing the DBMS Query Optimizer: Cardinality Estimation, Cost Model, and Plan Enumeration. Data Sci. Eng. 6(1): 86-101 (2021)
7. **Hai Lan**, Ke Han, Li Shen, Qiu Cui, Yuwei Peng. Accessing optimization with multiple indexes in TiDB. Journal of Computer Applications. 40(2): 410-515 (2020)
8. Song Wang, Yuwei Peng, **Hai Lan**, Qianwen Luo, Zhiyong Peng. Survey and Prospect: Data Integration Techniques. Journal of Software. 31(3):893-908 (2020)
9. **Hai Lan**, Song Wang, Yuwei Peng. A heterogeneous in-situ data querying system. National Database Conference (NDBC) 2019 demo paper.

Papers Under Review

10. **Hai Lan**, Zhifeng Bao, Shane Culpepper. Are We Ready for Updatable Learned Index in a Disk-Resident DBMS? - Analyses, Evaluations, and Enhancements. Submitted to VLDB 2023. (CORE A*)
11. **Hai Lan**, Yuanjia Zhang, Zhifeng Bao, Yu Dong, Dongxu Huang, Liu Tang, Jian Zhang. Why Query Plans are Different: An Automatic Detection and Inference System. Submitted to CIKM 2022. (CORE A)
12. **Hai Lan**, Yuanjia Zhang, Zhifeng Bao, Yu Dong, Dongxu Huang, Liu Tang, Jian Zhang. AutoDI: Towards an Automatic Plan Regression Analysis. Submitted to VLDB 2022 Demo Track. (CORE A*)

Google Scholar Page: https://scholar.google.com/citations?hl=en&user=N9aGCT4AAAAJ

HORNORS & AWARDS

- ✧ RMIT Research Stipend Scholarship (May. 2021)
- ✧ Second-Class Scholarship, Wuhan University (Dec. 2019)
- ✧ Second prize (2/70), TiDB[2] Hackathon 2018 (Dec. 2018)
- ✧ 2017 CCF National Excellent College Student (Only 104 students awarded across all universities in China) (Oct. 2017)
- ✧ Second-Class Scholarship, Wuhan University (Oct. 2016)
- ✧ National Scholarship (2/220), Wuhan University (Oct. 2015)
- ✧ First-Class Scholarship (10/220), Wuhan University (Oct. 2014)

RESEARCH EXPERIENCE

- ✧ Query Optimizer (Jun. 2021 – present)
  - Design an automatic tool to analyze the plan regression; work with TiDB optimizer team to enhance Cascades planner of TiDB.
- ✧ Trajectory Management System (Aug. 2020 – Sep. 2021)
  - Design the new storage schema for trajectory data and processing algorithms, including 8 typical trajectory queries, based on the new storage schema on the cloud setting; (2) implement the idea proposed in a real system.
  - Publish one paper (**publication #1**)

[2] A distributed HTAP database is designed and developed by PingCAP. It is inspired by Google F1 and Spanner. It has 31.3k stars at github. It is adopted by almost 1,000 companies around world.

- ✧ Index Recommendation based on Deep Reinforcement Learning (Aug. 2019 – Jun. 2020)
  - Design several rules to generate right index candidates. Design the suitable reward, state, action for index recommendation, and construct a DQN model for index recommendation.
  - Publish one paper **(publication #5)**
- ✧ Concept Prerequisite Relations for Massive Open Online Courses (Jun. 2021 – Sep. 2021)
  - Design a weaky supervised method to capture the concept prerequisite relations from the online courses' resources.
  - Publish one paper **(publication #2)**
- ✧ Improve the access performance in TiDB with DNF predicts. (Nov. 2018 – Nov. 2019)
  - Propose a new access method in TiDB to improve TiDB's performance when predicts are in DNF form.
  - Implement the method above and program has been merged into TiDB master.
  - Publish one paper **(publication #7)** as a result of the above research.
- ✧ Data Integration Techniques (May. – Aug. 2019)
  - Investigate the history of data integration and summarize a chronicle about techniques applied to data integration.
  - Publish one paper **(publication #8)**
- ✧ Heterogeneous in-situ data query (Dec. 2017 – Nov. 2019)
  - Design schema mapping between non-relational schema and relational schema. Design a query system with parser, optimizer, executor, metadata management.
  - Build a real system prototype which can use a query to access five different databases, i.e., relational, graph, key-value, time-series, non-structure database systems.
  - Publish one demo paper **(publication #9)**.
- ✧ Watermarking (May. 2016 – Dec. 2017)
  - Propose a reversible fragile watermarking scheme for the content authentication of spatial data, which embeds (detects) four kinds of watermarks into (from) each object (vertex).
  - Propose a multiple purpose watermarking scheme where robust watermark and fragile watermarks are embedded into the host map simultaneously.
  - Publish two papers **(publication # 3 & #4)** as a result of the above research

RESEARCH PROFILE

Hai's research interest is mainly on core database system problems, especially on query optimizer, indexing techniques, database tuning, spatial data processing. He has published several papers in VLDB, CIKM, and DSE. Hai has published one paper on index recommendation with deep reinforcement learning. This work is still one of state-of-the-art solutions in this field. A survey paper on query optimizer techniques has been conducted by Hai. He revisited the key problems in query optimizer and summarized all traditional methods and recent learning-based solutions, which gives the researchers and practitioners a clear view of solutions for query optimizer. Hai has worked with Alibaba spatial team and designed a comprehensive trajectory processing system and publish one paper in VLDB 2022. The proposed system is the first trajectory processing system that supports 8 typical trajectory queries. Hai also works with TiDB (one of the most popular distributed database systems) query optimizer team to enhance Cascades planner of TiDB, and also seeks to help optimizer developers quickly find the reasons of plan regressions. A tool on plan regression analysis has been deployed in TiDB.