

Xi Liang

xiliang.me | xiliang678@gmail.com | Google Scholar | Bay Area, CA

EDUCATION

University of Chicago

Chicago, IL

Ph.D. in Computer Science. Advisor: Prof. Sanjay Krishnan

2016.09 – 2022.03

Design and build efficient data systems with a specialty in approximate query processing

Ph.D. Thesis: Synopses for Efficient and Reliable Approximate Query Processing

EXPERIENCES

Senior Software Engineer | Databricks Inc. | Bay Area, CA

2022.03 – Present

- I designed and implemented new optimization strategies to the optimizer of Spark to improve the query performance.
- I contributed to the Enzyme which is a core component for MVs to achieve incremental view maintenance.
- I designed and implemented core components of the following framework/systems as part of Databricks runtime.
- Dreamcatcher: onling A/B testing framework for query replay to improve the stability and reliability of the Databricks runtime.
- Darwin: automatic online query performance regression detection and mitigation.
- Expression stats: collect and use file-level stats for expressions to improve data skipping.

Research Assistant | University of Chicago | Chicago, IL

2016.09 – 2022.03

- I designed and implemented the following research prototypes which lead to multiple first-author papers accepted by top conferences like SIGMOD.
- **JanusAQP**: Built on Apache Kafka and based on a novel dynamic partition tree that extends our previous work PASS, JanusAQP is a dynamic AQP system that can process insertions and deletions of data efficiently while provide query result with better accucflow latency.
- **PASS**: A system that combines offline-aggregates with stratified sampling for Approximate Query Processing. PASS features a partitioning algorithm that can minimize the worst-case error and a unique design that enables sample skipping to achieve a higher accuracy without sacrificing the latency.
- **PC**: A missing data analysis framework based on our novel synopsis called predicate-constraints. PC not only make it possible to encode beliefs and assumptions of missing data in a logical and formal way but also can be used to derive tight hard bounds for aggregate queries to achieve optimal reliability.
- **DQM**: A system that manages materialized views with deep reinforcement learning. Integrated with Apache SparkSQL, DQM demonstrates that robust view management to reduce the latency over evolving workloads can be achieved using reinforcement learning.

Software Engineer | Aurender Inc. | South Korea

2009.11 – 2013.08

- As one of the three software engineers the company started with, I worked on the award-winning music server “Aurender” until the company was acquired.
- As the only software engineer in charge of the server-side. I designed and implemented core features of the system from scratch. Including music library management; customized hardware support and control; remote-upgrading for shipped products; GUIs of the music player; RESTful APIs to interact with the iOS app, etc.
(C++/Python/QT/SQLite/LinuxDeviceDriver)

PATENTS

Compile time processing of extract, transform, load process.

Michael Armbrust, Yannis Papakonstantinou, ..., **Xi Liang**, et al. .

US Patent 12,517,905 (2026)

Incrementalization with Microtechniques Architecture.

Michael Armbrust, Yannis Papakonstantinou, ..., **Xi Liang**, et al. .

US Patent 12,386,833 (2025)

PUBLICATIONS

1. JanusAQP: Efficient Partition Tree Maintenance for Dynamic Approximate Query Processing.
Xi Liang, Stavros Sintos, Sanjay Krishnan.
IEEE 39th International Conference on Data Engineering, ICDE 2023.
2. Making Data Engineering Declarative.
Michael Armbrust, Yingyi Bu, Yannis Papakonstantinou, ..., **Xi Liang**, et al.,
Conference on Innovative Data Systems Research, CIDR 2023.
3. Sensor fusion on the edge: initial experiments in the EdgeServe system.
T Shaowang, **Xi Liang**, Sanjay Krishnan.
Proceedings of The International Workshop on Big Data in Emergent Distributed Environments, BiDEDE 2022.
4. Combining Aggregation and Sampling (Nearly) Optimally for Approximate Query Processing.
Xi Liang, Stavros Sintos, Zechao Shang, Sanjay Krishnan.
Proceedings of the 2021 International Conference on Management of Data, SIGMOD 2021.
5. CIAO: An Optimization Framework for Client-Assisted Data Loading.
Cong Ding, Dixin Tang, **Xi Liang**, Aaron J. Elmore, Sanjay Krishnan.
IEEE 37th International Conference on Data Engineering, ICDE 2021.
6. Fast and Reliable Missing Data Contingency Analysis with Predicate-Constraints.
Xi Liang, Zechao Shang, Aaron J. Elmore, Sanjay Krishnan, Michael Franklin.
Proceedings of the 2020 International Conference on Management of Data, SIGMOD 2020.
7. CrocodileDB: Efficient Database Execution through Intelligent Deferment.
Zechao Shang, **Xi Liang**, D Tang, C Ding, Aaron J. Elmore, Sanjay Krishnan, MJ Franklin.
Conference on Innovative Data Systems Research, CIDR 2019.
8. Opportunistic View Materialization with Deep Reinforcement Learning.
Xi Liang, Aaron J. Elmore and Sanjay Krishnan. arXiv:1903.01363, 2019.
9. Cross-layer refresh mitigation for efficient and reliable DRAM systems: A comparative study.
Xiaoan Ding, **Xi Liang**, Yanjing Li
IEEE International Test Conference, ITC 2017
10. Solving hex openings using job-level UCT search.
Xi Liang, I-Chen Wu.
International Computer Games Association, ICGA Journal 2015.
11. Job-level UCT search for solving Hex.
Xi Liang, I-Chen Wu.
IEEE Conference on Computational Intelligence and Games, CIG 2015.

PROFESSIONAL SERVICES

I served as a reviewer for the conferences and journals listed below.

- The 39th IEEE International Conference on Data Engineering (ICDE 2023).
- The 43rd International Conference on Distributed Computing Systems (ICDCS 2023).
- IEEE Transactions on Computers.
- Journal of Parallel and Distributed Computing.
- Journal of Computing.

AWARDS

2017, 2020: University Unrestricted Fellowship, University of Chicago.