

Yihao Hu

PhD Student in Computer Science

Department of Computer Science

Duke University

☎ (+1) 617-943-8060

✉ yihao.hu@duke.edu

🌐 <https://users.cs.duke.edu/~yh218/>

🐙 Github

Research Interest

I am broadly interested in query processing/optimization and data-intensive systems. My current research focuses on leveraging various techniques to build scalable systems that help users (1) analyze and understand how database queries are processed and executed, (2) formulate semantically correct queries to produce expected outputs.

Education

2020–present **Ph.D. in Computer Science**, *Duke University*, Durham, NC.

- Advisor: Prof. Jun Yang

2018–2020 **M.S. in Electrical & Computer Engineering**, *Duke University*, Durham, NC.

2014–2018 **B.S. in Computer Engineering & Electrical Engineering**, *Boston University*, Boston, MA.

- Advisor: Prof. Ari Trachtenberg, Prof. Alan Pisano

Research Experience

2020–present **Research Assistant**, *Duke Database Research Group*, Supervisors: Jun Yang, Sudeepa Roy, Kristin Stephens-Martinez.

- **HNRQ: Helping Novice Learn and Debug Relational Queries**

- Design and develop an SQL semantic debugger for helping users understand the execution of SQL queries by visualizing the SQL queries. The debugger and its underlying algorithms are designed to be generic and work with any database instance of various scales.

- **Explain Semantic Difference Between SQL Queries**

- Build frameworks that explain why two SQL queries are semantically different and propose small edits to one query to achieve query equivalence, by leveraging techniques from formal reasoning, Boolean algebra, etc.

Summer 2023 **Research Intern**, *ByteDance Infrastructure System Lab*, Supervisor: Yonghua Ding.

- **Adaptive Encoding for Columnar Store**

- Implemented state-of-the-art integer encoding algorithms within ByteHTAP distributive database system. Investigated potential directions for constructing encoding algorithms that are adaptive to the data pattern in columnar store using rule-based and machine learning methods.

Summer 2019 **Research Intern**, *Huawei Cloud CTO Office*, Supervisor: Tao Sun.

- **Compression and Deduplication of Log Data in NoSQL Database:**

- Designed and implemented a variation of gzip algorithm that compressed log entries with dynamic format. Implemented an in-memory cache policy which prevented log records with more than 90% similarity from being stored separately in the NoSQL database.

2017–2018 **Research Assistant**, *Boston University N.I.S. Lab*, Supervisor: Ari Trachtenberg.

- **Efficient Data Encryption for Collaborative Web App**

- Designed and implemented an encryption layer over Google Docs, which used a strong substitution cipher algorithm to dynamically encrypt the user-entered plaintext and decrypt the server-stored ciphertext for privacy protection.

Professional Experience

Summer 2020 **Software Engineer Intern**, *Amazon*.

- Root Cause Analysis (RCA) Tool for Amazon Dash Cart
 - Full-stack development on the RCA tool via AWS, which visualized the shopping data (add/drop item, product purchased, etc.) of each customer on a per-receipt basis. The tool was used by customer service/dash cart developers to review if the dash carts accurately capture all items purchased.

Publications

Peer-reviewed Full Paper

- 2024 **(Submission In Progress) Yihao Hu**, Zhengjie Miao, Kristin Stephens-Martinez, Sudeepa Roy, and Jun Yang. I-rex: A scalable debugger for sql queries. 2024.
- 2024 **(SIGMOD 2024 under review) Yihao Hu**, Amir Gilad, Kristin Stephens-Martinez, Sudeepa Roy, and Jun Yang. Qr-hint: Actionable hints and edits for query debugging. 2024.
- 2019 **Yihao Hu**, Ari Trachtenberg, and Prakash Ishwar. Collaborative privacy for web applications. In *2019 57th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 460–469, 2019.

Peer-reviewed Demo Paper

- 2022 **Yihao Hu**, Zhengjie Miao, Zhiming Leong, Haechan Lim, Zachary Zheng, Sudeepa Roy, Kristin Stephens-Martinez, and Jun Yang. I-rex: An interactive relational query debugger for sql. In *Proceedings of the 53rd ACM Technical Symposium on Computer Science Education V. 2, SIGCSE 2022*, page 1180, New York, NY, USA, 2022. Association for Computing Machinery.
- 2020 Amir Gilad, **Yihao Hu**, Daniel Deutch, and Sudeepa Roy. Muse: Multiple deletion semantics for data repair. *Proc. VLDB Endow.*, volume 13, page 2921–2924. VLDB Endowment, Sep. 2020.

Awards

- 2017 UROP Student Research Award, Boston University

Teaching & Mentoring

- Fall 2021 Teaching Assistant, Introduction to Database Systems (CompSci 316), Duke University
- Summer 2021 Student Mentor, Duke CS+: CompSci Projects Beyond the Classroom
- Fall 2020 Teaching Assistant, Introduction to Database Systems (CompSci 316), Duke University
- Fall 2019 Teaching Assistant, Computer & Information Security (ECE 590), Duke University
- Fall 2017 Teaching Assistant, Cybersecurity (EC 521), Boston University

Skills

- Programming C/C++, Python, JavaScript, Java
- Databases PostgreSQL, MySQL, CockroachDB, MongoDB
- Web Dev Flask, Django, AngularJS, Vue, React
- Tools Docker, PyTorch, AWS, Google Cloud