Jaehyun Ha

Department of Artificial Intelligence, POSTECH, South Korea
Homepage: rntlqvnf.github.io
Advisor: Wook-Shin Han

Data Systems Lab @ POSTECH
Email: jhha@dblab.postech.ac.kr

Research Interests

Building AI-native database systems for unified analytics and inference on multi-modal data

- **Graph Databases:** Architecting a foundational query framework based on a unified graph representation of multi-modal data
- **Semantic Operator:** Embedding Large Language Models (LLMs) into the database to enable context-aware querying of unstructured data
- **Query Optimization:** Designing proxy models and novel cardinality estimation methods to optimize the execution of computationally expensive semantic operators

Education

POSTECH, Graduate School of Artificial Intelligence — Mar. 2022 – Present
M.S./Ph.D. Integrated Program
Data Systems Lab
Advisor: Prof. Wook-Shin Han

POSTECH, Dept. of Computer Science and Engineering — Mar. 2018 – Mar. 2022
B.S. in Computer Science and Engineering
GPA: 4.06 / 4.3
Class of 2021: Graduated with the highest GPA in CS Dept.

Ulsan Science High School — Mar. 2016 – Nov. 2017

Publications

Graph Systems

VLDB'26 Lee, T., **Ha, J.**, Tak, B., Han, W. S.
Submitted *S62: Schemaless Graph Engine Strikes Back for General-Purpose Analytics.*

SIGMOD'24 Lee, W. (equal contribution), **Ha, J. (equal contribution)**, Han, W. S., Park, C., Park, M., Han, J., Lee, J.
DoppelGanger++: Towards Fast Dependency Graph Generation for Database Replay.

VLDB'24 Lee, W., **Ha, J.**, Han, W. S., Park, C., Park, M., Han, J.
Demo *DoppelGanger++ in Action: A Database Replay System with Fast Dependency Graph Generation.*

Query Optimization

PODS'23 Kim, K., **Ha, J.**, Fletcher, G., Han, W. S.
Guaranteeing the O(AGM/OUT) runtime for uniform sampling and size estimation over joins.

Projects

Semantic Operator Optimization — Jun 2025 – Sep 2025
Visiting Scholar Research at UIUC with Prof. Yongjoo Park

- **Problem:** Semantic operators are prohibitively expensive in latency and dollar cost. Existing optimization methods (i.e., proxy models) present a poor trade-off, forcing a choice between unacceptably low accuracy or still-significant latency and cost, which limits practical adoption
- **Contribution:** Developed novel methodologies to discover and build high-performance proxies that significantly improve latency and dollar cost while preserving accuracy
- **Outcome:** Plan to submit to SIGMOD 2026

High-Performance Schemaless Graph Database System Jan 2023 – Jun 2025

- **Problem:** The conversion of unstructured data into knowledge graphs produces schemaless graphs, where nodes and edges have their own different schemas. However, existing database systems have significant limitations in performing high-performance analytics on such data
- **Contribution:** Designed and implemented a full-stack system with a specialized storage, optimizer, and query engine tailored for schemaless graph analytics
- **Role:** Core developer for a large-scale system (246K+ LOC); authored approx 50% of commits
- **GitHub:** https://github.com/postech-dblab-iitp/turbograph-v3
- **Outcome:** Submitted to VLDB 2026

High-Speed Dependency Graph Generation for Database Replay Dec 2022 – Dec 2023

Industry Collaboration with SAP Labs Korea

- **Problem:** Database replay systems are innovative tools for capturing and replaying real-world workloads for testing purposes. However, their real-world efficiency is severely bottlenecked by the slow process of dependency graph generation
- **Contribution:** Proposed an efficient algorithm to accelerate dependency graph generation
- **Impact:** Implemented on the commercial SAP HANA system, demonstrating real-world utility
- **Outcome:** Accepted at SIGMOD 2024 and VLDB 2024 (Demo)

Theoretically Optimal Join Cardinality Estimation Algorithm Mar 2022 – Nov 2022

- **Problem:** Existing sampling-based algorithms for join cardinality estimation lacked a provable optimal O(AGM/OUT) runtime bound
- **Contribution:** Developed the first algorithm to achieve this theoretical optimal bound
- **Outcome:** Accepted at PODS 2023

Professional Experiences

University of Illinois Urbana-Champaign (UIUC), IL, USA Jun 2025 – Sep 2025

Visiting Scholar

- Performed research on the optimization of semantic operators
- Conducted a joint research project under the supervision of Prof. Yongjoo Park

Digital Platform team, SK hynix Inc., Korea Jul 2020 – Aug 2020

Software Engineer Intern

- Developed real-time video comment overlay system for SK hynix's internal video streaming systems
- Won silver prize (2nd) in internship contest

Honors and Fellowships

- POSTECHIAN Fellowship 2022,2023
- Samsung Dream Scholarship Foundation Scholar 2019-2021

Teaching Experiences

- **Teaching Assistant**, AIGS540: Big Data Processing Spring 2024
- **Teaching Assistant**, CSED421: Database Systems Spring 2022