

Jaehyun Ha

Department of Artificial Intelligence, POSTECH, South Korea
Homepage: github.com/jhha
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 M.S./Ph.D. Integrated Program <i>Data Systems Lab</i> Advisor: Prof. Wook-Shin Han	Mar. 2022 – Present (Expected: Feb. 2028)
---	--

POSTECH , Dept. of Computer Science and Engineering B.S. in Computer Science and Engineering GPA: 4.06 / 4.3 <i>Graduated with the highest GPA in CS Dept.</i>	Mar. 2018 – Mar. 2022
---	-----------------------

Ulsan Science High School	Mar. 2016 – Mar. 2018
----------------------------------	-----------------------

Professional Experiences

Google Systems Research Group , Sunnyvale, CA, USA <i>Visiting Student Researcher</i> ◦ Researching hardware-software co-optimization to accelerate semantic queries ◦ Advised by Yeounoh Chung and Helena Caminal	Jun 2026 – Sep 2026
--	---------------------

University of Illinois Urbana-Champaign (UIUC) , IL, USA <i>Visiting Scholar</i> ◦ Conducted research on semantic operator optimization, resulting in a paper accepted at SIGMOD 2027 ◦ Joint research project under the supervision of Prof. Yongjoo Park	Jun 2025 – Sep 2025
--	---------------------

Digital Platform team , SK hynix Inc., Korea <i>Software Engineer Intern</i> ◦ Developed real-time video comment overlay system for SK hynix's internal video streaming systems ◦ Won silver prize (2nd) in internship contest	Jul 2020 – Aug 2020
--	---------------------

Publications

1. **Jaehyun Ha**, Yongjoo Park, Wook-Shin Han. *CADENZA: Compiling Natural-Language Intent into Task-Specific Operator DAGs for Semantic Query Processing*. SIGMOD 2027.
2. **Jaehyun Ha**, Yongjoo Park, Wook-Shin Han. *CADENZA in Action: Breaking the Monolith with Intent-Dependent Plan Spaces for Semantic Queries*. VLDB 2026 (Demo).
3. Taesung Lee, **Jaehyun Ha**, Byungchul Tak, Wook-Shin Han. *TurboLynx: Schemaless Graph Engine Strikes Back for General-Purpose Analytics*. VLDB 2026.
4. Taesung Lee, **Jaehyun Ha**, Byungchul Tak, Wook-Shin Han. *TurboLynx in Action: A Schemaless Graph Engine for General-Purpose Analytics*. VLDB 2026 (Demo).
5. Wonseok Lee*, **Jaehyun Ha***, Wook-Shin Han, Changgyoo Park, Myunggon Park, Juhyeong Han, Juchang Lee. (*equal contribution) *DoppelGanger++: Towards Fast Dependency Graph Generation for Database Replay*. SIGMOD 2024.
6. Wonseok Lee, **Jaehyun Ha**, Wook-Shin Han, Changgyoo Park, Myunggon Park, Juhyeong Han. *DoppelGanger++ in Action: A Database Replay System with Fast Dependency Graph Generation*. VLDB 2024 (Demo).
7. Kyoungmin Kim, **Jaehyun Ha**, George Fletcher, Wook-Shin Han. *Guaranteeing the $O(AGM/OUT)$ runtime for uniform sampling and size estimation over joins*. PODS 2023.

Invited Talks

- Semantic Operator Optimization, Google Systems Research Group Jan 2026

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:** Accepted at SIGMOD 2027 and VLDB 2026 (Demo)

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:** Accepted at VLDB 2026 and VLDB 2026 (Demo)

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 $O(\text{AGM}/\text{OUT})$ runtime bound
- **Contribution:** Developed the first algorithm to achieve this theoretical optimal bound
- **Outcome:** Accepted at PODS 2023

Technical Skills

Programming Languages	C/C++ (Advanced), Python (Advanced), SQL (Advanced), CUDA (Intermediate), Java (Intermediate), JavaScript (Intermediate)
Databases & Data	PostgreSQL, OracleDB, DuckDB, Kuzu, Neo4j, pandas
Systems & Build Tools	Linux, Git, Docker, Kubernetes, CMake, Bazel, uv
ML/LLM	PyTorch, Hugging Face, OpenAI API, Azure OpenAI API

Honors and Fellowships

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

Teaching Experiences

- | | |
|---|-------------|
| ◦ Teaching Assistant , Agentic RAG, Samsung Electronics | Spring 2026 |
| ◦ Teaching Assistant , AIGS540: Big Data Processing, POSTECH | Spring 2024 |
| ◦ Teaching Assistant , CSED421: Database Systems, POSTECH | Spring 2022 |

References

Wook-Shin Han

wshan@dblab.postech.ac.kr

POSTECH Distinguished Professor, Pohang University of Science and Technology (POSTECH), Korea

Yongjoo Park

yongjoo@g.illinois.edu

Assistant Professor, Department of Computer Science, University of Illinois Urbana-Champaign (UIUC), USA