

Yihao Liu

✉ yhliu918@gmail.com ☎ +86 18810097393 🔗 <https://yhliu918.github.io/> 📄 [yhliu918](#)

Education

- Peking University** Yuanpei College *Sept 2017 – Jul 2021*
Bachelor in Data Science
- **GPA:** 3.7/4.0 (Rank: 15%)
 - **Advisor:** Bin Cui
- Tsinghua University** Institute for Interdisciplinary Information Sciences(IIS) *Sept 2021 – Now*
PH.D. in Computer Science
- **research direction:** Cloud database system / In-memory execution engine / Compression techniques.
 - **Advisor:** Huanchen Zhang

Internship

- Recommendation System Algorithm Engineer Intern** *ByteDance, Beijing*
ByteDance *Sept 2020 – Jan 2021*
- Cold-start advertisement placement strategy exploration: proposed a strategy of “maximizing information gain” to help the system converge to an optimal strategy quickly.
- Software Engineer Intern** *DP Technology (start-up)*
DP Technology *Mar 2021 – Aug 2021*
- Contributor of **deepmd-kitv2.0**: improved the embedding process in the Deepmd-kit molecular simulation algorithm. <https://github.com/deepmodeling/deepmd-kit>
- AI Research Intern** *Shanghai AI Lab*
Shanghai AI Lab *May 2025 – Now*
- Diffusion Language model inference acceleration.

Publications

- Selective Late Materialization** VLDB 2025
Yihao Liu, Shaoxuan Tang, Yulong Hui, Huanchen Zhang
<https://github.com/yhliu918/duckdb/tree/latest> 🔗
- The concept and trade-offs of late materialization (LM) vary greatly in modern vectorized databases compared to that in traditional column stores. Aggressively applying LM strategies no longer yields optimal results. Instead, we propose a selective LM approach that strategically applies materialization based on attributes and materialization points. After being integrated into DuckDB, this approach achieves an average latency improvement of 16% on the join order benchmark, demonstrating its effectiveness in accelerating real-world queries.
- LeCo: Lightweight Compression via Learning Serial Correlations** SIGMOD 2024
Yihao Liu, Xinyu Zeng, Huanchen Zhang
<https://dl.acm.org/doi/10.1145/3639320> 🔗
- Introduce LeCo (i.e., a Learned Compression framework), that achieves a Pareto improvement on both compression ratio and random access speed among lightweight compression techniques. LeCo enables a 5.2x speed up in the Arrow columnar execution engine and a 16% increase in RocksDB’s throughput.
- Cost-Intelligent Data Analytics in the Cloud** CIDR 2024
 Huanchen Zhang, **Yihao Liu**, Jiaqi Yan
<https://www.cidrdb.org/cidr2024/papers/p78-zhang.pdf> 🔗
- OkraLong: A Flexible Retrieval-Augmented Framework for Long-Text Question Answering** EMNLP 2025 Under Revision
 Yulong Hui, **Yihao Liu**, Yao Lu, Huanchen Zhang
<https://arxiv.org/abs/2503.02603> 🔗

* means equal contribution.

Projects

Embryo: a high-performance distributed analytical database

Now

- Main contributor to the Embryo database, developing the core logic of the execution engine and network communication between workers.

A book-sharing website with a distributed database backend

2021

- Built a book-sharing website with a distributed NoSQL database as its backend. The database uses MongoDB, Mongos, Redis, and HDFS as building blocks.

Distributed system (MIT 6.824)

2022

- Built a fault-tolerant distributed key-value store with linearizability.

Docking scoring function in Drug discovery

rf_vina_enhance 

- Introduced a new scoring function of docking, leveraging non-bonded interaction information and outperformed VinaRF20.

Technologies

Languages: C++, Python, SQL, Java

Systems worked on: RocksDB, Parquet, Arrow, Presto(Velox), DuckDB

Service

SIGMOD 2024 ARI (Availability & Reproducibility Initiative) committee