LeFeng Wang

(+86) 15611094516 | wanglefeng@sjtu.edu.cn | lfwang-sjtu

Summary

Research Interests My research interests lie in applying verification techniques to computer systems, ensuring correctness and efficiency in complex software infrastructures. In particular, I focus on database management systems (DBMS), applying formal verification to enhance query optimization, improving execution efficiency while ensuring semantic correctness.

Highlight 5 years of programming experience (primarily in C/C++, with additional proficiency in Python, Java, and JavaScript). 1 year of hands-on development experience in complex open-source projects, including work on **openGauss** , a system built on a PostgreSQL-based architecture. 1 year of research experience on an academic project, fostering strong collaboration, critical thinking, and problem-solving skills.

Education

Shanghai Jiao Tong University *Shanghai, China*

B.Eng in Software Engineering *Sep. 2020 - Jun. 2025 (expected)*

- Overall GPA: 79.76/100
- Average GPA from Fall 2020 to Fall 2023: 74.69/100
- Average GPA from Spring 2024 to present: **90.06/100**
- Relevant Courses: Architecture of Applications (86), Development of Internet Application (94), Intelligent information system modeling (89), Science and Technology Innovation (92), Unmanned System (95), Data Mining and Big data anlysis (87.5), Design and Implementation of data intelligence platform (88), Design and development of Internet products (88), Design and development of Internet platform and maintenance (88).

Publication

[1] **GRewriter: Practical Query Rewriting with Automatic Rule Set Expansion in GaussDB [Industry]**

Zhe Jiang, Zhaoguo Wang, Haoning Lan, Chuzhe Tang, Haoran Ding, **Lefeng Wang**, Songyun Zou, Zhuoran Wei, Yongcun Liu, Xiang Yu, Yang Ren, Guoliang Li, Haibo Chen.

To be submitted to VLDB 2025.

Research Experience

Shanghai Jiao Tong University *Shanghai, China*

Research intern at IPADS, advised by Prof. ZHAOGUO WANG *Apr. 2024 - Nov. 2024*

- **Research Topics:** Query Rewriting, Rule-based Optimization, Formal Verification
- Proposed an expressive and enumerable rule language
- Extended the capabilities of existing rule enumerators **WeTune** to support new rule language
- Implemented a new query rewrite engine in **openGauss** that can load or unload rewrite rules at runtime.
- Evaluated the new query rewriting engine on datasets such as TPC-H and TPC-DS, as well as real-world applications, including a banking transaction system and Huawei's internal ERP system. The engine accelerated more than half of the SQL queries by over 3.6×, with the fastest query achieving a 1543× speedup.

Honors & Awards

2021 **Undergraduate Excellence Scholarship**, top 10% students in SJTU. *Shanghai, China*

2020 **Zhiyuan Honorary Scholarship**, top 5% students in SJTU. *Shanghai, China*

Project Portfolio (Selected)

Query Rewrite Rule Enumerator

Developer

- Developed a SQL query rewrite rule enumerator based on the **[WeTune]** enumerator.
- Extended WeTune's capabilities to support enumerating rules written in a more expressive rule language.
- Applied several innovative pruning strategies to enhance the efficiency of the existing automatic rule discovery, reducing the enumeration process that would have taken the original enumerator years to just a few weeks.
- Integrated an existing SQL verifier into the system by converting each rule into a set of SQL queries through an algorithm that is rigorously formally verified, ensuring the correctness of the rewrite rules.
- Optimized more than half of the real-world SQL queries by over 3.6× using the rules discovered through this enumerator.

LSM-KV

Developer [code]

- Designed and implemented a key-value store in C++ based on the Log-structured Merge Tree (LSM-Tree), leveraging key-value separation to optimize storage efficiency and read performance.
- Implemented garbage collection using file hole-punching techniques to efficiently reclaim storage space.
- Optimized SSTable reads with Bloom filters to improve query performance.

Online Bookstore

Developer [backend] [frontend]

- Developed a full-stack web application, covering the design of the database(MySQL), backend(Spring Boot), and frontend(React).
- Integreted multiple databases, including MySQL, MongoDB and Neo4j to support diverse data storage and usage requirements.
- Designed the system with a microservices architecture to enhance performance and scalability.
- Integrated a message broker(Kafka) for order management and transaction control to ensure robust consistency and reliability.

Consignment Platform

Team Leader & Developer [code]

- A financial product consignment platform that manages transactions between customers and companies. We have creatively designed a framework that allows platform managers to assemble business operations by combining predefined atomic services through visual orchestration.
- By closely collaborating with my teammates, we were able to work extremely efficiently and completed the entire project in less than two weeks, leveraging the resource management platform provided by IST lab.

Big Data Analytics Service

Team Leader & Developer [code]

- A data service designed to analyze the PTMTorrent dataset to extract valuable insights into pretrained models. To efficiently handle the large-scale dataset, we utilized Apache Storm for real-time stream processing and Apache Hive for storing and querying the data. The service runs on a self-configured cluster, ensuring scalability and high performance.

MapReduce Simulator

Developer [code]

- A C++ project that simulates the operation of MapReduce using threads to emulate a coordinator server and worker servers. The program runs on an inode-based filesystem, also developed as part of this project, which is structured into three layers: block layer, inode layer, and filesystem layer.

Skills

Programming C/C++, Python, Java, JavaScript, SQL, Matlab

Frameworks and Tools Linux, Git, Spring Boot, Hadoop, Apache Storm, Kafka, PyTorch.

Databases (used in projects): MySQL, PostgreSQL, MongoDB, Neo4j, Redis