

Chi Zhang

iskyzh@gmail.com | github.com/skyzh | skyzh.dev | linkedin.com/in/alex-chi-skyzh

Education

Carnegie Mellon University

2022/08 – 2023/12

Master of Science in Computer Science, GPA 4.10/4.33

Pittsburgh, PA, USA

- Teaching Assistant for 15-445/645 Database Systems (Fall 2022, Spring 2023, Fall 2023)
- Courses: Distributed Systems, Compiler Design, Advanced Database Systems, Deep Learning Systems, etc.

Shanghai Jiao Tong University

2018/09 – 2022/06

Bachelor of Engineering in Computer Science and Technology

Shanghai, China

- GPA 93.80/100, Rank 1/149, National Scholarship 2019 (Top 0.2% national-wide)
- A+ Courses: Operating Systems, Computer Architecture, Computer Networks, and 28 others

Work Experience

Neon

Since 2024/02

Systems Software Engineer

Remote / Pittsburgh, PA, USA

... and also

2023/05 – 2023/08

Software Engineer Intern

Remote / Pittsburgh, PA, USA

- Neon is a fully-managed PostgreSQL service built on a key-value storage engine with point-to-time recovery support.
- **Compaction Strategy Enhancement.** Conducted an in-depth analysis and evaluation of the storage engine to assess performance metrics and storage space efficiency. Implemented the RocksDB-style tiered compaction and improved page reconstruction strategy, which reduced space amplification by 2x and enhanced read-update performance by 20%.
- **Improved User Adoption on the Edge.** Enhanced the overall reliability of the Neon serverless driver and the control-plane proxy. Collaborated closely with the [Prisma](#) ORM team to integrate the serverless driver into Prisma and ensured compatibility with Vercel Edge Runtime by transitioning the Rust Prisma engine codebase to be WebAssembly-ready.

RisingWave Labs

2021/08 – 2022/07

Database System R&D Intern

Shanghai, China

- **Top Contributor of [RisingWave](#).** RisingWave is a database system with PostgreSQL-compatible interface that incrementally maintains materialized views. Worked on features including streaming index joins, query optimization of stream plans, distributed streaming execution, cloud-native LSM state store, vectorized expression framework.
- **Streaming Index Joins.** Designed shared state and streaming index in RisingWave; implemented index lookup join executor; implemented delta join DAG optimizer transformations; implemented distributed delta join scheduler.
- **Performance Improvement.** Conducted intensive benchmarks and analyzed performance issues. Fixed bugs, proposed strategies, and led cross-team collaboration which improved the system throughput by 10x in a 3-month period.
- **Developer Experience.** Initiated the RiseDev development tool and the developer dashboard, which is deeply integrated into the development workflow across debugging, unit testing, integration testing, and benchmarking.
- **Mentoring.** Mentored database kernel interns and helped their successful integration into the team. Maintained overview documents of the database kernel to facilitate knowledge transfer and help new hires learn about the system.

ByteDance

2021/06 – 2021/08

Storage System R&D Intern, TerarkDB Team

Beijing, China

- **Co-Optimized [TerarkDB](#) and [ZenFS](#).** TerarkDB is a fork of RocksDB and ZenFS is a filesystem on Zoned Namespaces (ZNS) SSDs. Implemented Zone-aware Garbage Collection in TerarkDB for ZNS and WAL-Aware Zone Allocator in ZenFS, which reduced 3-4x of space amplification and greatly improved tail latencies caused by zone allocation.

PingCAP

2020/08 – 2021/01

Storage System R&D Intern, TiKV Storage Team

Shanghai, China

- **Built LSM-based storage engine [AgateDB](#) from ground-up.** Inspired by WiscKey and BadgerDB, AgateDB separates large values from the LSM tree into a separate value log, so as to reduce write amplification and improve throughput.

... continues on the next page ...

Open-Source Contributions

BusTub github.com/cmu-db/bustub as *Teaching Assistant for Database Systems* 2022/08 – 2023/12

- Lead the development of the BusTub educational database system and course projects in CMU Database Systems course.
- Added query processing layer to the system with PostgreSQL syntax support. Restructured the query execution project.
- Added multi-version concurrency control to the system based on HyPer/Umbra undo log version chain implementation.
- Redesigned course projects to help students better understand the concepts and align with industrial database systems.
- Developed leaderboard tests to challenge advanced students and enable further study in optimizing database systems.

RisingLight Maintainer github.com/risinglight Since 2022/01

- Lead the development of RisingLight, an OLAP database system in Rust for educational purpose. RisingLight supports simple TPC-H queries, and has a merge-tree based columnar storage.

TiKV Maintainer github.com/tikv Since 2020/05

- Maintain TiKV Coprocessor, the push-down execution framework of TiDB. Served as mentor for the Linux Foundation LFX Mentorship. Mentored community members to contribute new features (e.g., enum data types, plugin system).

Personal Projects github.com/skyzh 6.6k followers on GitHub

- **mini-lsm** (2k Stars) Build a simple LSM-Tree storage system in Rust in a week.
- **type-exercise-in-rust** (1k Stars) Learn Rust generics by implementing a vectorized expression evaluation framework.

Research Experience

Adaptive Query Optimization Framework github.com/cmu-db/optd 2023/09 – 2023/12

CMU Database Group, advised by Professor Andy Pavlo Pittsburgh, PA, USA

- **Developed optd**, an optimizer framework based on the Columbia Cascades paper targeting real-time OLAP queries.
- **Adaptive Optimization.** optd collects statistics during execution and uses runtime data to guide later plan searches.
- **Partial Exploration.** optd explores plans by reusing and incrementally expanding the plan space from the last search.

PostgreSQL Extension Manager github.com/cmu-db/pgextmgrext 2023/02 – 2023/05

CMU Database Group, advised by Professor Andy Pavlo Pittsburgh, PA, USA

- **Implemented pgextmgrext**, a PostgreSQL extension that manages other PostgreSQL extensions and provides new APIs to PostgreSQL extension developers that enables them to write new extensions with fewer lines of code.
- **Integration with PostgreSQL Ecosystem.** Integrated `pg_hint_plan` with the extension manager. Implemented output rewriter in the extension manager, and based on that, a demo extension `pg_poop` that rewrites all text to poop emojis.

Skills

Programming Languages: Rust (6 years), C++, Python, Node.js and Golang

Tech Skills: Stream-Processing Systems, Database Systems (Optimizer and Query Execution), Key-Value Storage Systems, SSD-optimized File Systems

Last Updated in February, 2024