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

Databricks 2025/06 – Present

Member of Technical Staff, Neon Storage Team Pittsburgh, PA, USA

Neon, acquired by Databricks 2024/02 – 2025/06

Systems Software Engineer, Storage Team Pittsburgh, PA, USA

- Work in the Neon storage team with the focus on the fundamentals of the storage layer-map structure; implement new compaction strategies (e.g., gc-compaction) to optimize amplifications of the Neon pageserver service — a key-value storage system storing Postgres write-ahead-logs and pages with copy-on-write branches and point-in-time recovery support.
- Design, develop, and ship storage features to meet the product's requirements and customers' needs. Examples: store small files in the key-value storage system (for logical replication); non-continuous sparse key space; detach from parent branches.

Neon 2023/05 – 2023/08

Software Engineering Intern Pittsburgh, PA, USA

- **Compaction Enhancement**. Conducted an in-depth analysis and evaluation of the storage engine to assess performance metrics and storage space efficiency. Implemented a demo of RocksDB-style tiered compaction and improved page reconstruction strategy, which reduced space amplification by 2x and enhanced read-update performance by 20%.
- **Serverless SQL Driver**. Enhanced the overall reliability of the Neon serverless driver. Developed a Web-Assembly-based solution and collaborated closely with the Prisma ORM team to integrate the serverless driver into Prisma.

RisingWave Labs 2021/08 – 2022/07

Database System R&D Intern Shanghai, China

- **Top Contributor[1] of RisingWave**. RisingWave is a Postgres-compatible database with streaming processing support. As a founding member of the company, I worked on things like Rust rewrite of the codebase, streaming index joins, query optimizer for streaming, distributed stream processing, cloud-native LSM state store, and vectorized expression framework.

ByteDance, Storage System R&D Intern, TerarkDB Team Beijing, China, 2021/06 – 2021/08

- **Co-optimized TerarkDB and ZenFS**. TerarkDB is a fork of RocksDB and ZenFS is a filesystem on zoned namespaces SSDs (ZNS). Implemented zone-aware garbage collection in TerarkDB for ZNS and WAL-aware zone allocator in ZenFS; with both combined reduced 3-4x of space amplification and greatly improved tail latencies caused by zone allocation.

PingCAP, Storage System R&D Intern, TiKV Storage Team Shanghai, China, 2020/08 – 2021/01

- **Built LSM 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 ...

[1] in terms of number of pull requests merged to the main branch; until 6 months after I left the company

Open-Source Contributions

Personal Projects github.com/skyzh 7.7k GitHub followers

- **mini-lsm** (3.3k stars, top 20 on Hacker News) A self-guided course to build an LSM key-value storage engine in Rust in a week, and then extend its compaction algorithm and implement multi-version concurrency control in the next two weeks.
- **tiny-llm** (1.9k stars, top 10 on Hacker News) Serve the Qwen2 model on Apple Silicon from scratch only with matrix APIs. The course covers implementing the RoPE, FlashAttention, quantized matrix computations, continuous batching, etc.
- **type-exercise-in-rust** (1.4k stars) Learn Rust generics by implementing a vectorized expression evaluation framework.
- **write-you-a-vector-db** (710 stars) A course of turning any relational database into vectordb (based on CMU's BusTub).

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 Committer, 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).

Selected Blog Writings on skyzh.dev

- State Store in Streaming Systems, a survey of state store implementations of Apache Flink, Materialize, and RisingWave.
- Key-Value Separation in LSM Storage Engines, covering RocksDB's BlobDB, TerarkDB, BadgerDB, and TiKV's TitanDB.
- Lessons Learned from Building an Extensible Optimizer series, a collection of ideas and lessons from my optd experience.

Research Experience

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

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

- One of the main developers of CMU-DB's optd extensible optimizer framework. It is based on the Columbia/Cascades optimizer framework and supports 15 out of 22 TPC-H queries. Key features: partial exploration, adaptive query optimization, Apache Datafusion integration, subquery unnesting, physical properties, and async-task-based optimizer implementation.

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

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

- A PostgreSQL extension manager as an extension that manages other PostgreSQL extensions and provides new APIs to the PostgreSQL extension developers. One of the new APIs is the output rewriter, and based on that, I wrote a demo extension pg_poop that rewrites all query result into poop emojis.