Pei Chen

https://peichen-cs.github.io
+86 188 1152 5703 ⋄ chenp732@gmail.com

EDUCATION

Tsinghua University **Sept 2018 - Jan 2022**
MPhil of Computer Science and Technology Beijing, China
Advisor: Prof. Jiwu Shu and Prof. Youyou Lu, Storage Research Group
Seleted Courses: Advanced Operating System(4.0), High Performance Computing Experiment(4.0)

Central South University **Sept 2014 - June 2018**
Bachelor of Computer Science and Technology; **GPA: 4.00/4.00**; **RANK: 1**% Changsha, China
Selected Courses: Advanced Mathematics(4.0), Compiler(4.0), Operating Systems(4.0), Discrete Mathematics(4.0), Data Structure(4.0), Database(4.0), Algorithm Analysis And Design(4.0), Introduction to Parallel Algorithms(4.0), Distributed System(4.0), Computer Architecture(4.0), Network Engineering(4.0), Embedded System Design(4.0)

PUBLICATIONS AND PATENTS

SNlog: A SmartNIC-driven shared log.
Pei Chen, Youyou Lu, Qing Wang, Junru Li, Jiwu Shu. In Submission

Efficient and Consistent NVMM Cache for SSD-based File System.
Youmin Chen, Youyou Lu, **Pei Chen**, Jiwu Shu.
Submitted to IEEE Transactions on Computers. ***TC 2018***

A global address space management method for distributed persistent memory.
Jiwu Shu, Youmin Chen, Qing Wang, **Pei Chen**, Youyou Lu.
CN111241011A. *Published June 2020*

A memory communication method and device based on RDMA.
Youyou Lu, Jiwu Shu, Youmin Chen, **Pei Chen**, Jun Xu, Peng Lin.
CN111858418A. *Published December 2020*

Multi-read and multi-write log system for SmartNIC.
Jiwu Shu, Qing Wang, **Pei Chen**, Youyou Lu, Jianye Yao, Yue Zhao.
Filed April 2021. *Chinese National Patent Pending*

RESAERCH EXPERIENCE

SNlog: Multi-Reader and Multi-Writer Log System Based on SmartNIC Dec 2019 - July 2021
Project Leader *Advisor: Prof. Jiwu Shu, Storage Research Group, Tsinghua University*

- Proposed **SmartNIC for direct storage management**, eliminating file system overhead and bypassing usual limits of ARM processing power.
- Introduced efficient log handling, data separation, and fast data transfers with RDMA.
- **Quadrupled** performance without draining server resources.

DPMALLOC: Distributed Persistent Memory Address Management System Mar 2019 - Oct 2019
Project Leader *Advisor: Prof. Jiwu Shu, Storage Research Group, Tsinghua University*

- Developed a reliable and high-speed memory allocator for distributed systems. Introduced a state machine in the allocator and a special protocol for address allocation without locks.
- Filed a patent, ID 201911418599.X.

HGDSM: CPU/GPU Distributed Shared Persistent Memory System July 2018 - Jan 2019
Lead Researcher *Advisor: Prof. Jiwu Shu, Storage Research Group, Tsinghua University*

- Developed an RDMA communication technique with GPUDirect to minimize message copying.
- Grouped and rated the network connections, so as to balance the saturation and thrashing of the cache space.
- Improved write speed by **18**. Filed a patent with ID CN111858418.

AFCM: Efficient and Consistent NVMM Cache for SSD-based File System Jan 2018 - May 2018
Researcher *Advisor: Prof. Jiwu Shu, Storage Research Group, Tsinghua University*

· Introduced Adaptive Fine-grained Cache Management (AFCM) for persistent memory, merging pages and cache lines. This reduces issues from both page and cache line-only approaches.
· Implemented a Transaction Copy-on-Write (TCOW) strategy for data safety.
· **83%** higher throughput than SCCM. Published in Transactions on Computers 2018.

INDUSTRY EXPERIENCE

Task-based execution framework - scaling out for GaussDB Oct 2024 - present
Huawei Gauss *code lines: 6K — Beijing, China*

· Implemented logical connection based on the socket for reuse. In larger-scale scenarios, it can enhance operations without being constrained by the number of socket connections. Implementing primary-standby SMP allows leveraging the standby machine's CPU to enhance performance in read scenarios.
· Implemented stateless thread pool where all state resides within self-contained logical threads. Binding and unbinding involve only lightweight pointer switching, supporting SQL-level task switching. This resolves the issue of overly coarse transaction-level scheduling in thread pools, which often leads to thread pool saturation and inability to release resources.
· The framework also incorporates a Morsel-driven execution mode. This alleviates the limitations of pre-execution optimizer planning and post-execution cgroup control, which previously failed to effectively mitigate the impact of slow/poor SQL queries on the entire system due to coarse-grained control. The degree of parallelism and resources can elastically change during query execution, so the dispatcher can react to execution speed of different morsels but also adjust resources dynamically in response to newly arriving queries in the workload.

Operator Parallelization Support Based on SMP Oct 2023 - Oct 2024
Huawei Gauss *Beijing, China*

· Implemented Operator Parallelization based on SMP, including insert into select, merge into, update, delete, and upsert, resulting in the OPS improvement of 5x.
· Implemented Parallel Shared Hash Join based on SMP, resulting in the OPS improvement of approximately 2x.
· Providing plan caches for Q Message, same as P/B/E Messages, resulting in the decrease of hard parse during execution.

Columnar Analytical Engine Based on Object Storage Service July 2023 - Sep 2023
Huawei Cloud *Beijing, China*

· Created a Java MergeEngine focusing on object metadata and version tracking.
· Removed secondary index for object metadata and improved bandwidth of index service for object storage by 50%.
· Reduced the cycle of AP operations from 1 day to 15 minutes.

Scalability and Reliability Optimization of Cloud Storage Metadata Oct 2022 - June 2023
Huawei Cloud *Beijing, China*

· Developed a MongoDB-based routing strategy, decreasing metadata access failure time from 30 minutes to 30 s.
· Optimized route scalability with MongoDB, cutting down access latency from 10 seconds to 0.6 milliseconds.
· Enhanced MongoDB chunk assignment, improving failure response from 96% to 99.99%.

Optimization of Local Storage Read Performance for Cloud Storage Metadata May 2022 - Sep 2022
Huawei Cloud *Beijing, China*

· Enhanced point-lookup with a hash index, reducing seek time by 21.8%, boosting throughput by 10%.
· Conducted basic consistency checks on the LSM-tree.

Introducing vector engine of ClickHouse into MySQL Jan 2022 - April 2022
ByteDance *Beijing, China*

· Created a C++ MysqlExecutor inspired by ClickHouse's ScanExecutor to work with ByteNDB storage.
· Presented the updated schema and identified C++ classes for modification to support ByteNDB storage.

HONORS AND AWARDS

Outstanding Undergraduate Thesis Award(Top 2%) of Central South University *2018*
Outstanding Graduate(Top 0.1%) of Hunan Province & Central South University *2018*
Honorable Mention of the International Mathematical Contest in Modeling(MCM) *2016*
National Scholarship(Top 0.2% Nationwide) *2016*
First-class Scholarship(Top 1%) *2016*
Qu Yuan Scholarship(Top 0.1%) *2016*
National Encouragement Scholarship(Top 5%) *2015*
Second-class Scholarship(Top 5%) *2015*

SERVICE AND MEMBERSHIP

- Shadow Reviewer, Eurosys 2026
- Lecturer, Tsinghua University Student Career Development Association Student Tutor Group 2022 - 2023
- Guest speaker, Tsinghua University Computer Science Department "Future of Computing" Phd and Master's Forum 2021

SKILLS AND INTERESTS

Programming Languages: C, C++, Shell, Python, LaTeX, Java, Assembly(x86), Go
Parallel Computing Skills: MPI, Linux perf, CUDA, OpenMP
Languages: Chinese (Native), English (Fluent)