

刘芷溢

📞 13470385770 | ✉️ lzy_CS_LN@163.com | 🏠 个人网站

教育经历

- 电子科技大学 计算机科学与技术 2024.9 - 至今
硕士 GPA: 3.6/4.00 四川, 成都
- 电子科技大学 网络空间安全 2020.9 - 2024.6
学士 GPA: 3.81/4.00 四川, 成都

项目经历

- **Oceanbase 数据库大赛 2024** 2024.10 - 2024.12

技术栈: C++, 数据库, 分布式系统

- 初赛是在开源 MiniOB 数据库上完成 25 个功能扩展并支持向量功能, 决赛是对 Oceanbase 的向量索引查询性能进行优化。
- 初赛: 功能扩展包括语法解析、查询优化、逻辑算子、物理算子编写等, 个人主要实现了增加 text 类型、复杂子查询、增加视图、别名以及向量索引查询优化等操作, 加深了对通用数据库工作流程的理解, 最终以满分成绩获得省赛第一名。
- 决赛: 优化 ann_benchmarks 基准测试和混合标量后过滤的性能。涵盖工程优化与算法优化。
 1. 对 HNSW 索引的参数调优, 将基础测试 QPS 从 592 提升到 1095。
 2. 对回表流程进行优化, 减少磁盘 IO 访问次数, 混合标量查询延迟降低 30
 3. 重构向量索引关键字段, 并对混合标量类查询进行谓词下推, 混合标量查询延迟降低 60
 4. 对数据进行 int8 量化并增加 vsag 库接口, 减少 HSNW 向量相似度计算的耗时。
 5. 其他优化包括 simd 指令优化、peo 算法移植等, 最终在基础测试中 QPS 性能提升 201

- **TinyKV: 一个具有高可用性、可扩展性的分布式 KV 数据库** 2025.4 - 2025.6

技术栈: Go, 分布式系统, 数据库

- 参考 etcd, 实现 Raft 共识协议, 包括领导者选举、日志复制、领导者变更、日志压缩快照以及集群内单个成员的变化。参考 etcd 的测试测试 Raft 实现的正确性。
- 基于实现的 Raft 协议, 结合基于 LSM-Tree 的 badgerDB 实现高可用的存储引擎。
- 实现多版本并发控制 (MVCC), 通过时间戳序列化数据版本, 基于 Percolator 两阶段提交模型构建分布式事务框架的存储引擎部分。

- **Bustub: 一个支持简单 SQL 语句的单机数据库** 2025.2 - 2025.4

技术栈: C++, 计算机网络, 数据库

- 使用哈希表和双向链表实现 LRU-K 算法, 基于该算法实现线程安全的 Buffer Pool; 同时在内存页和磁盘页的换入以及访问时加锁, 为页面管理提供线程安全的 RAII 风格守卫操作。
- 使用 latch crabbing 算法实现并发的 B+ 树索引, 同时应用乐观锁优化, 大幅提升写性能。
- 采用火山模型 (Volcano Execution Model) 实现查询执行引擎, 完整支持:
 - * 数据操作: SELECT、DELETE、UPDATE
 - * 高级功能: JOIN、AGGREGATION、LIMIT、DISTINCT、External Merge Sort
 - * 实现双模式连接算子: 适用于小数据集关联的嵌套循环连接/嵌套索引连接 (Nested Loop Join/Nested Index Join) 和优化大规模数据匹配的哈希连接 (Hash Join)。
- 为事务实现多版本并发控制 (MVCC) 和可串行化验证 (OCC), 基于 MVCC 重构插入、删除、更新、顺序扫描以及索引扫描算子, 支持并发多版本下的索引更新和数据更新, 支持事务 abort 后数据恢复; 使用 OCC 向后验证来实现可串行化隔离级别。

获奖经历

- **Oceanbase 数据库大赛决赛 11 名** 2024.12

全国大学生系统能力大赛

- **优秀奖学金 (一等)** 2024-2025

电子科技大学

技能和兴趣

- 编程语言: C++, rust, Python
- 技术: 数据库, 计算机网络, 分布式系统, 操作系统
- 兴趣: 数据库, 分布式系统, 篮球