

# 王亮

☎ 187-7460-8517 | ✉ iggiawang@gmail.com | 🌐 iggiawang.cn | 📍 hey-kong

## 教育

华中科技大学, 计算机系统结构, 学硕

2021 年 9 月 – 2024 年 6 月 (预期)

- 研究方向: 海量存储系统与云存储服务

武汉大学, 软件工程, 学士

2017 年 9 月 – 2021 年 6 月

- GPA: 3.83/4.0, 排名: 16/258
- 主修课程: 数据结构与算法 (99)、数据库系统 (93)、操作系统原理 (86)、网络及分布式计算 (92)

## 经历

华为技术有限公司 (Huawei), 深圳

2020 年 11 月 – 2021 年 4 月

Cloud BU 基础服务产品部实习生

- 参与分布式文件系统 SwarmFS 的开发。实现了 Posix 语义的 lookup 与 read 流程; 使用内存池对消息 msg、事务 trans 等对象进行管理; 统计 RocksDB 读写链路上各个过程的具体时延; 参考分布式 ID 生成器 Leaf, 实现并发创建文件时无阻塞地生成 Inode ID; 编写单元测试用例 (gtest), 进行性能测试 (vdbench、mdtest)。

北京平凯星辰科技发展有限公司 (PingCAP), 北京

2019 年 7 月 – 2019 年 8 月

PingCAP Talent Plan 第三期线下学员

- 在 TiDB 上开发 Index Advisor, 设计索引推荐算法, 实现与 Client 交互。算法主要受 DB2 Advisor 论文启发, 利用 TiDB 基于代价的优化器以及 What-If 能力为输入的 SQL 推荐最佳索引。用 TPC-DS 测试, 推荐的索引创建后能有效缩减 SQL 执行时间, 少数 SQL 能获得 20 倍左右的性能提升。

## 项目

🔗 **LanguorDB**: 一个用 Go 语言实现的 LevelDB 基础版, 并做了优化。设计了一种惰性 Compaction 策略, 牺牲读性能来避免频繁触发归并操作, 减少写放大, 提高写性能; 同时进一步实现了并行查找机制和 Row Cache 两种方法来优化读性能。对 4GB 数据集进行随机读写性能测试, LanguorDB 的写时延比 LevelDB 减少 27%, 平均查询时延比 LevelDB 减少 58%。

🔗 **MayflyCache**: 模仿 groupcache 实现的一个分布式缓存。基于 LRU 缓存淘汰策略, 使用一致性哈希算法选择节点, 实现负载均衡, 节点之间采用 HTTP 协议进行通信, 使用 protobuf 序列化数据进行传输, 优化了节点间通信的性能, 同时还实现了并发访问控制机制, 防止缓存击穿。

🔗 **hbtrees**: 基于 DRAM-NVM 的混合索引存储系统, 内存部分为学习索引 ALEX(SIGMOD '20), NVM 部分为 B+Tree。设计了一种冷热分离架构, 一个 ALEX 叶子节点对应一棵 B+Tree, 根据热度将 ALEX 叶子节点分为热节点和冷节点。对于热节点, 写数据时先写日志, 然后更新 ALEX 叶子节点中的缓存, 后台线程根据日志将数据持久化至 B+Tree, 读数据时直接访问 ALEX 叶子节点中的缓存; 对于冷节点, 读写都需要访问 NVM 中的 B+Tree。

🔗 **城市房价数据分析系统**: 基于 Flask 和 Vue.js 前后端分离的团队项目, 获得了武汉大学计算机学院 2020 年暑期实训项目一等奖、最佳技术创新奖。主要负责房价数据爬取、数据库设计以及后端开发。

## 技能

- **Programming Languages**: 主要使用 Go, 另外用过 C++/Rust/Python
- **Technical Skills**: 熟悉数据库 (TiDB, MySQL)、键值存储系统 (RocksDB, WiscKey, Bourbon, etcd)、分布式相关知识, 熟练使用 Linux, 了解 Kubernetes/RDMA/gRPC 等技术

## 其他

- 英语水平良好 (四级 573、六级 504)
- 完成过 🔗 MIT 6.824、🔗 CSAPP 等国外计算机课程 Lab
- 热爱开源, 为多个知名开源项目贡献过 PR, 是 TiDB Active Contributor (contributed  $\geq 8$  PRs)