

# 柳湛宇

手机: (+86) 18888912197 · 邮箱: zhanyuliu@zju.edu.cn

求职岗位: 后端开发工程师/数据库开发工程师



## 教育背景

- 浙江大学 计算机科学与技术学院 软件工程(非软院) 硕士 2022.09-至今
- 浙江大学 竺可桢学院 & 计算机科学与技术学院 软件工程(非软院) 本科 2018.09-2022.06
- GPA: 3.89/4
  - 竞赛与奖项: 浙江省政府奖学金、浙江大学二等奖学金、浙江大学竺可桢学院“领航”奖学金、浙江大学计算机学院“卓越工程师”奖学金、全国高校计算机能力挑战赛程序设计赛全国一等奖、浙江省“挑战杯”大学生课外学术科技作品竞赛省赛一等奖、浙江省物理创新竞赛二等奖等。

## 实习经历

- 字节跳动 国际化电商-数据智能中台 后端开发实习生 2024.03-2024.07
- 业务简介: 针对多个业务域实体(用户、商品、达人等)的标签属性提供圈选、明细查询、趋势洞察、多形式数据消费与精细化运营 ROI 分析等能力;
  - 参与梳理 [TB/天] 级别的数据链路产出与优化方案, 参与开发、上线了核心底表数据产出链路由原生 HiveQL 数据任务到 spark 任务的迁移, 上线后平稳运行且单节点任务执行耗时大幅缩小至 20%-50%。
  - 参与开发实现“算法动态人群”模块, 对接资源位和算法工程两个模块实现了基于算法排序的动态广告资源投放。项目目前在 ROW 和 EU 上线, 实验阶段 CTR、GPM 整体上平均有 30% 幅度的提升。
  - 参与分群实时判定与 Diff 消息产出架构升级, 参与了分群圈选规则的优化和 RocksDB 数据底表的存储优化改造, 添加一系列数据产出监控。架构部分上线后提供了对账能力和有效监控能力。
- 浙江智奥科技有限公司 (DolphinDB) 数据库内核研发实习生 2023.11-2024.03
- 优化 DolphinDB 时序数据存储底层 LSM 引擎, 新增高写入场景的多版本 SST 文件合并策略。经过测试, 在高写入场景下, 可以获得读写事务综合 13%+ 的执行速度提升。
  - 设计实现了数据和 redolog 日志存储多 Volume 配置、数据再平衡和写入分流策略功能。实现了存储空间的热扩展, 且在纯 HDD、高写入的场景, 有 5%+ 的执行速度提升。
  - 为 DolphinDB 新增了一系列滑动窗口型计算函数, 并将部分内核计算函数扩展为支持高可用流计算引擎的算子函数;
- 菜鸟网络科技有限公司 国内供应链技术部 Java 开发实习生 2021.05-2021.09
- 负责菜鸟内部仓储管理系统(WMS)的维护与功能开发。主要参与开发了“云鹏”订制品仓配一体化项目, 独立实现了大部分数据链路管理逻辑, 以及订制品打印流程的可靠控制逻辑, 项目上线后平稳运行;
  - 初步熟悉了阿里内部一系列中间件的基本使用和日志追踪、持续集成等非代码软件平台的使用; 初步了解了仓储管理系统相关的业务流程。

## 开源项目

- 石原子科技有限公司 (StoneDB) StoneDB 研发实习生 2023.06-2023.08
- 修改 MySQL 语法解析文件以及对应的引擎接口, 实现可自定义的 SQL 语句形式的指标查询和参数配置功能。减少了 StoneDB 语法的 shift/reduce 冲突数并从系统表层扩展了引擎配置方法。这一系列工作已于2023年开源之夏项目中作为进阶项目结项;
  - 参与 StoneDB 商业版系统中向量化执行引擎的部分算子的开发、测试和优化; 参与其列式存储引擎的原生分布式方案初步设计与实现。

## 项目经历

### 内存数据库 Cubebase

- 项目简介: 对标 Aerospike, 哈希分区的原生分布式混合存储 KV 数据库。是目前我导师所在公司自研流计算平台的存储层。参与开发任务主要包括:

- 构建支持定制化索引读写的谓词索引：扩展 FishStore 的思路实现谓词辅助索引，显著减少扫全表次数，良好支撑特化查询。正则匹配查询速度总体上提升 200%+ 且转化为 1 篇专利；
- 基于热度感知的无锁动态桶扩展算法：将原有固定桶数量的哈希索引修改为根据访问热度使用 CAS 无锁操作进行动态扩容缩容的哈希索引。100% 读性能相比旧设计提高 20 %；
- 混合存储模式实现基于共享内存的集群节点快速热重启：使用混合存储模式时通过将磁盘分区的主索引和辅助索引构建在共享内存中，使得重启时不需要解析持久化存储的数据即可重建索引。可以实现 TB 级数据的秒级重启。

## 分布式 MiniSQL

- 简介：基于已有的单机 SQL 引擎，构建主从架构的分布式数据库；通过 Zookeeper 作为分布式协调中心，netty 作为 IO 框架，实现了节点选举、数据划分、日志复制、集群容错、负载均衡等功能的分布式数据库。

## 用户态线程库

- 简介：使用 C 语言与内嵌汇编编写通过保存现场与切换用户态执行栈实现的用户态线程库，实现了以 Work Stealing 为核心的类似 Golang 的 GMP 调度方法。经过测试，在大规模并发场景显著优于 pthread。

## 学术论文

### Discovering the Most Influential Geo-Social Object Using Location Based Social Network Data

(第二作者)

2020-10

- 解析 LBSN 数据展示的异构图，并构建对应数据结构，从社交网络图谱关系、文本内容相似度、地理位置邻近程度三个方面构建评价函数，进而计算出具有影响力的节点，实现影响力最大化

## 个人能力

- 编程语言：有一定的 C++、Java、Golang 语言编程经验；
- 编程知识与技能：
  - 对数据结构与算法、计算机组成原理、计算机网络、操作系统等基础知识有一定程度掌握；
  - 熟悉 Java 语言基础知识，对 JVM 虚拟机执行逻辑和 Java 多线程模型及其原生/响应式多线程编程框架有一定了解；对 Spring 框架及其实现原理有一定了解，有以 SpringBoot+Vue 为基础框架的前后端分离应用程序开发经验；
  - 学习过 15.445、6.824 等经典的数据库系统课程。对以 MySQL/PostgreSQL 为代表的关系型数据库的处理逻辑、存储结构、索引构建等底层原理有一定程度的了解；对基于 LSM 树结构的 KV 数据库底层原理有一定了解；对基于 Raft 共识算法有一定的了解；
  - 有一定 linux 环境下 C++ 的开发经验，对 linux 命令有一定了解；对 GNU/linux 环境下如 cgroup、共享内存与信号量、epoll 等特性工具和框架有一定了解；
  - 对 Kafka、Redis 等常用中间件有一定基础了解和使用经验；
  - 能熟练使用 git，掌握 maven、cmake 等代码构建工具的基础使用。
- 语言 CET6 565

## 附录

- 项目地址：
  - 分布式 MiniSQL: <https://gitee.com/StreakyPork/distributed-mini-sql>
  - Bustub 数据库: <https://github.com/streakyPorker/bustub2022>
  - 用户态线程库: <https://gitee.com/StreakyPork/coroutine-lzy>