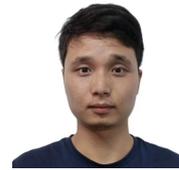


胡进磊



📍 华科计算机硕博连读 | 📅 预计 2027 年 6 月毕业
☎ (+86) 131-2500-7880 | ✉ hujinlei1999@qq.com | 🌐 SimpleHappy2027

个人总结

- 主要研究方向为：**高性能存储系统**。学生一作发表 CCF A 类期刊 1 篇，CCF B 类论文 3 篇（其中一篇入选 Best Paper Candidates）。
 - 硕士阶段：基于持久内存 (PM) 的高性能索引结构。
 - 博士阶段：基于新型闪存设备 (ZNS SSD) 的键值存储系统。
- 熟悉 C/C++ 编程与 Linux 底层 I/O 栈，熟悉单机存储引擎与硬件特性 (Hash/LSM-Tree/B+-Tree/SSD/PM)，熟悉使用 LLM 工具 (Codex/Gemini/ChatGPT)，了解向量检索系统 (参与相关论文选题以及竞赛)。

教育经历

华中科技大学，计算机科学与技术学院，计算机科学与技术，博士	2023.09 - 至今
信息存储及应用实验室，导师：冯丹教授	
华中科技大学，武汉光电国家研究中心，计算机系统结构，硕士	2020.09 - 2023.06
电子科技大学，计算机科学与工程学院，信息安全，本科	2016.09 - 2020.06

主要研究内容

博士研究课题：面向新型闪存设备的键值存储系统性能优化研究 2023.09 - 至今

- 基于 ZNS SSD 的 B+ 树优化存储引擎 ZKV (CCF A 类期刊 TACO'26)
 - 针对现有 B+ 树型键值存储引擎存在的写入机制与新型分区接口特口不匹配问题，设计基于新型分区接口的 B+Tree 型高性能存储引擎，提供了高效的读写性能以及高闪存资源利用率。
- 基于 ZNS SSD 的 LSM-Tree 空间放大优化研究 (在投中)
 - 针对现有 LSM-Tree 型存储引擎在 ZNS SSD 上的空间放大问题，设计基于 SSTable 合并优化的空间管理机制，避免冗余写入并提高了空间利用率。
- 基于 ZNS SSD 的学习型索引闪存缓存系统 (在写中)
 - 针对现有闪存缓存系统的元数据开销、高垃圾回收开销、低命中率问题，设计基于学习型索引与准入算法结合的新型闪存缓存系统，充分提高闪存缓存系统的命中率、使用寿命以及使用生命周期。

硕士研究课题：基于持久内存的高性能索引结构 2020.09 - 2023.06

- 针对扩容操作优化的无锁哈希结构 R2Hash (CCF B 类会议 ICCD' 25 Best Paper Candidates)
 - 针对现有键值存储在内存缓存层的低效数据布局问题，设计了基于混合内存的新型扩容友好的无锁哈希结构，提供低查询时延以及高写入带宽。
- 针对结构化操作和空间优化的高性能 B+ 树索引 SSTree (CCF B 类会议 ICCD 2024)
 - 针对现有基于持久内存的 B+ 树索引存在的结构化 (节点分裂) 操作开销大、空间利用率低等问题，采用链表拓展形式的叶子节点以及异步写入机制，提升持久 B+ 树索引的读写性能。
- 路径压缩优化的读写友好的 radix 基数树索引 Rwort (CCF B 类会议 ICCD 2023)
 - 通过分层压缩机制，显著降低了树的高度。集成了微型布隆过滤器，以减少的不必要访问。通过惰性分割标志和双向链接指针来减少关键路径延迟。

向量检索相关研究与竞赛

- 基于闪存的图向量索引布局优化机制 Giant (CCF B 类 DASFAA'26 录用，共同一作)
 - 背景挑战：针对现有基于 SSD 的图向量检索系统 (如 Starling 等) 在处理高维数据时，由于页面可容纳节点数骤减导致的扩展候选不足，以及单一广度扩展造成的 I/O 时间利用率低下等性能瓶颈。
 - 核心方法：设计并实现了一种新型的页面分组扩展机制，充分重叠 I/O 等待与计算时间：

- * **无映射开销的常驻内存分组索引**: 将节点进行空间分组并建立局部 Vamana 索引; 通过磁盘布局与 PQ 编码重排, 彻底消除了节点到页面的映射表内存开销, 实现更深层的候选节点探索。
- * **页面扩展列表**: 充分利用 SSD 页面尾部对齐的空闲空间, 存储经过多样性剪枝的高质量补充候选节点, 在零额外 I/O 成本下丰富了候选池。

- **实验结果**: 相比现有 SOTA 方法 (Starling, DiskANN), 在同等召回率下, 平均检索延迟显著降低 **12%-43%**, 系统吞吐量 (QPS) 提升 $1.10 \times - 1.41 \times$ 。

全国大学生计算机系统能力大赛——第 2 届 PolarDB 数据库创新设计赛 2025
(外卡优胜奖第 14 名, 共 3500 余支队伍)

- - 针对 pgvector 插件原生 HNSW 索引构建时间长、索引性能低问题, 根据分析工具 (火焰图、埋点) 定位构建、查询过程性能瓶颈 (距离计算、读取 buffer)。
- 通过 RabitQ 量化、SIMD 距离加速计算、将邻居的量化码保存在同一页面等优化方法, 在 Text2Img@0.85 数据集上, 将 QPS 查询性能提高一倍 (3078 \rightarrow 6093), 构建时间降低 30% (780s \rightarrow 503s)。
- **AI 应用与异构内存编程挑战赛暨首届全国人工智能创新应用大赛** 2021
(第三名奖金 3 万)
 - 针对大规模浮点数据的向量检索加速问题, 设计了 DRAM/PM 混合内存加速方案, 相比 DRAM 版本索引空间提高 4 倍, 相比 PM 版本索引性能提高了 12 倍。
 - 关键技术: 基于 DRAM/PM 的分级缓存; 设计新式哈希结构; 硬件优化。

校企合作项目经历

全闪栈千亿级元数据管理算法优化 (某 IT 基础设施供应商 - 组长) 2021.10 - 2023.6

- **项目背景**: 针对全闪存阵列在管理千亿级超大规模元数据时面临的严重写放大、CPU 占用过高及持久化性能抖动问题, 研发支持新型存储介质 (NVM) 的高速存储引擎。
- **核心攻坚**: 设计并实现了一套“内存索引-缓存合并-闪存管理”三层去耦合架构:
 - **无锁化与碎片治理**: 在内存索引层设计 B-Tree 批量聚合算法, 将离散请求合并处理, 实现读写操作的无锁化, 有效缓解碎片化问题。
 - **持久化与平滑 I/O**: 在缓存合并层引入日志机制, 将随机写转换为大块顺序落盘; 同时结合 NVM 设计元数据缓存聚合机制, 平滑频繁更新带来的性能波动。
 - **高效空间管理**: 在底层构建混合索引结构, 实现空闲块的高效分段查找, 大幅降低长尾查询延迟。
- **项目成果**: 实现了高 I/O 吞吐下的低 CPU 占用率, 保障了存储引擎在大规模元数据高频交互场景下的长时间运行稳定性。

基于 NVM 的全闪阵列地址映射表缓存 (某网络基础服务商 - 主要研发核心) 2020.10 - 2021.06

- **项目背景**: 为突破大规模闪存阵列地址映射表的访存性能瓶颈, 利用 NVM 的非易失性及高随机读写特性, 设计具备极高并发能力与崩溃一致性的新型缓存系统。
- **核心攻坚**:
 - **高性能哈希引擎设计**: 实现读无锁与轻量级写, 优化哈希冲突避免机制以降低内存访存开销, 并支持桶级别的热度感知。
 - **写放大抑制与一致性保证**: 引入极少量 DRAM 作为热点数据缓冲, 有效解决 NVM 的写放大顽疾; 创新性地实现了“无日志冗余写”的数据崩溃一致性保证。
- **项目成果**: 主导完成了核心代码研发与多线程环境调优。经 YCSB 性能压测, 该系统在单机 36 线程下, 读/写吞吐量分别高达 **26 M IOPS** 与 **8 M IOPS**。

其他

- **兴趣爱好**: 羽毛球、健身、做饭、电影