



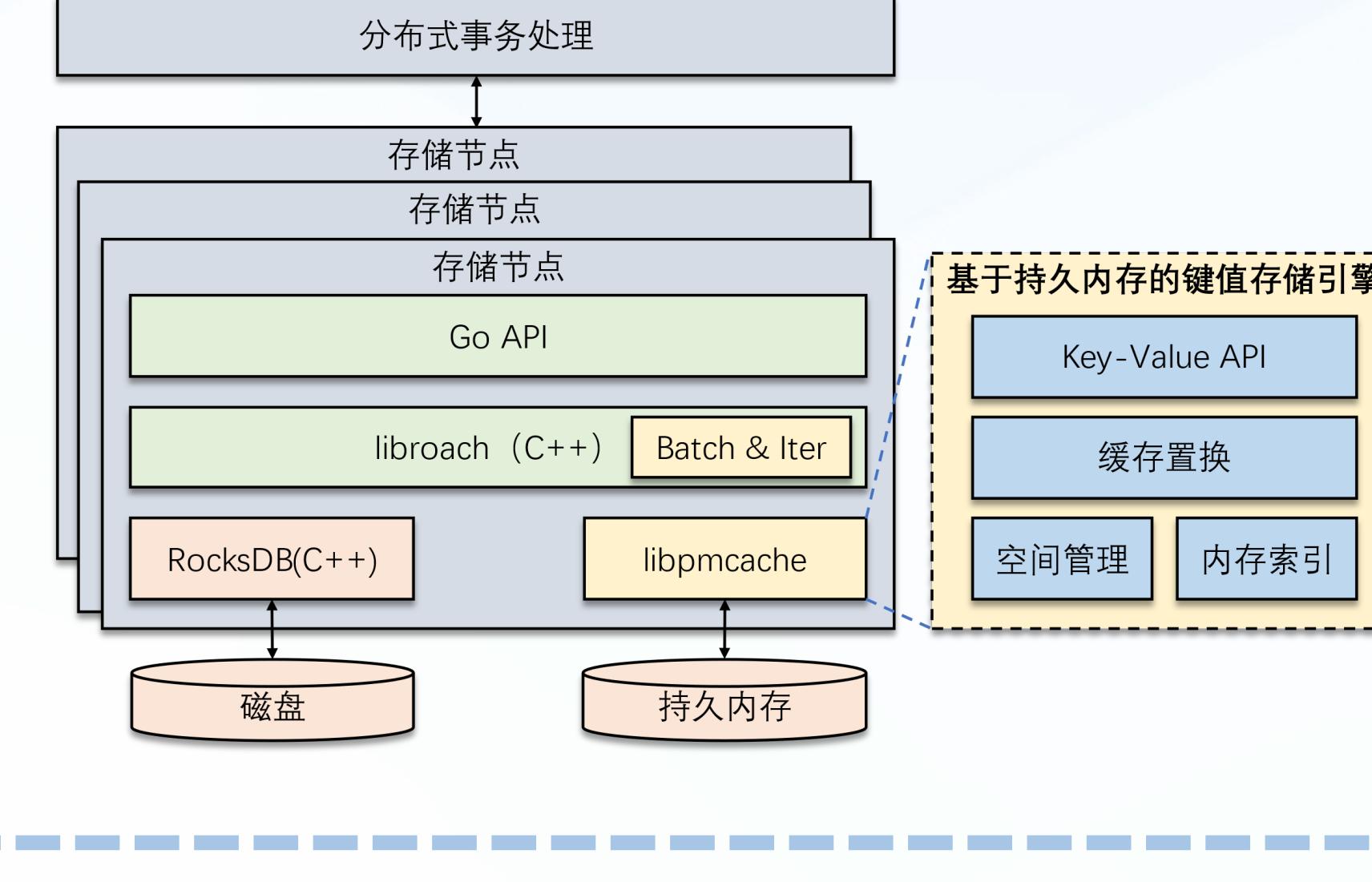
基于持久内存的数据库事务处理机制优化

唐震，李京龙，王伟

联系方式：唐震，15652917972，tangzhen12@otcaix.iscas.ac.cn

工作介绍

- 分布式数据库系统广泛用于支撑高并发事务场景
- 持久内存作为一种新型介质，用于优化数据库事务处理机制，可有效提升吞吐率并简化故障恢复流程
- 我们提出了基于持久内存的键值存储引擎，包括：
 - 改进的空间管理机制
 - 改进的支持范围检索的索引机制
- 基于CockroachDB实现原型系统，并验证了有效性



问题分析

事务执行过程中存在大量需要被持久化的中间数据，但在事务回滚或重试时会被丢弃，造成资源浪费。引入持久内存后可以暂存这些中间数据，提升性能。在此场景中，需应对如下挑战：

- 如何应对持久内存读写性能不平衡及写放大特性？
- 如何针对持久内存设计空间管理机制，以应对中间数据的大量小规模随机读写操作？
- 数据库并发控制机制中大量使用到范围查找，如何针对持久内存设计优化的索引机制？
- 如何确保数据正确写入持久内存？

技术细节

在存储空间管理机制方面：

- 固定大小且对齐的存储块及多个线程访问区
- 单个线程只能访问指定编号的线程访问区
- 使用改进的freelist组织空闲空间
- 首尾添加校验字段，保障原子操作的完整性



分块存储结构

在索引机制方面：

- 将索引全部置于内存 (DRAM) 中，节点失效后从持久内存重建索引
- 改进基数树索引，从key中首个相异字符位置开始匹配

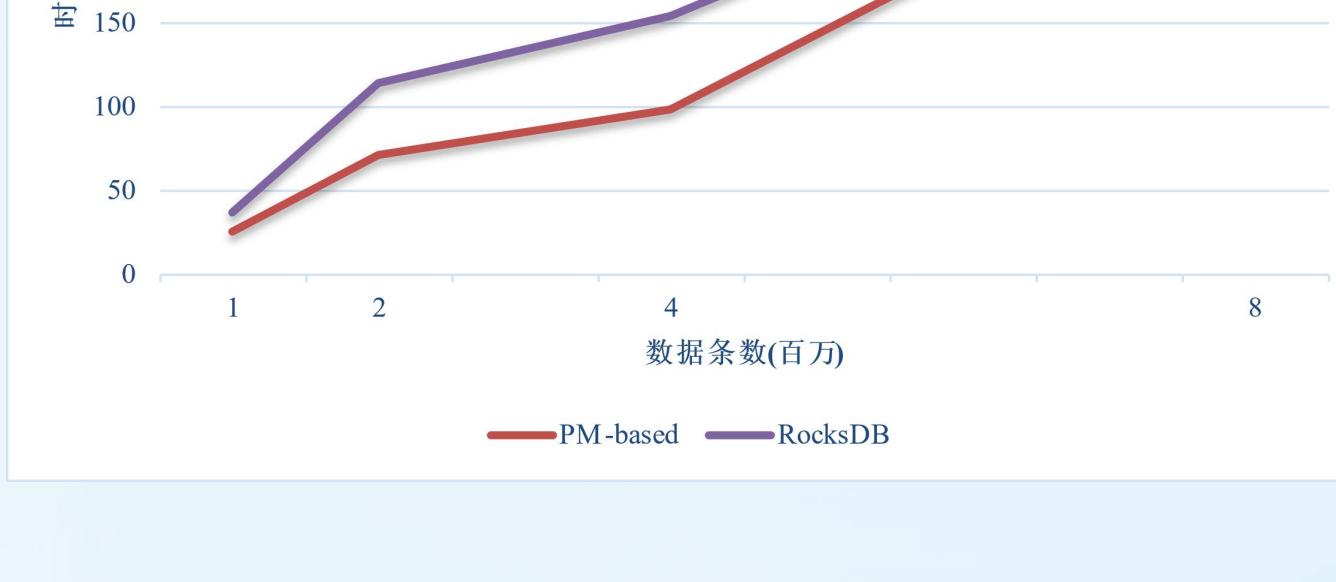


索引结构

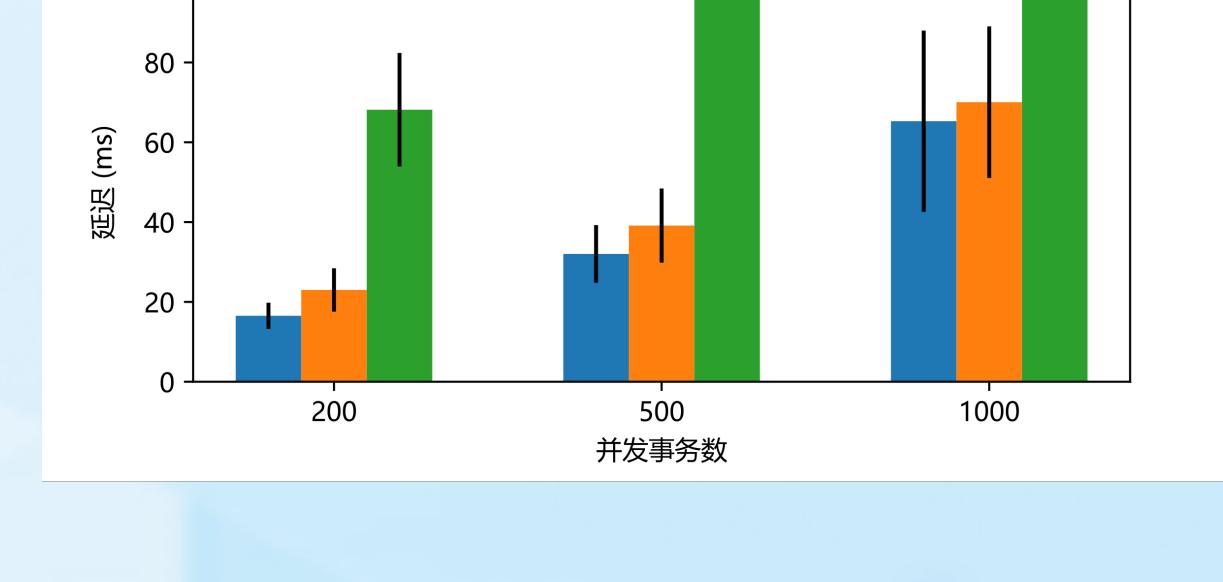
实验结果

基于CockroachDB实现原型系统，修改了CockroachDB与RocksDB交互的libroach模块。实验在配备Intel Optane持久内存的3台服务器上进行。

键值存储的读写性能测试
读、写、删除等操作延迟较原始系统至多降低50%



改进并发事务机制的性能测试
在高并发事务场景下，事务执行时间最多降低29%



TPC-C基准测试
平均事务执行时间降低11%
99百分位延迟至多降低55%。

