



中国科学院高能物理研究所
Institute of High Energy Physics
Chinese Academy of Sciences

十二月份考核报告

张 航 畅
导师：顾旻皓
计算机应用技术

2024. 12



截止到2024.08

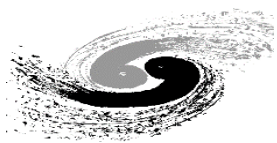
➤ 基于分布式内存缓存的在线计算方案

- ◆ 内存缓存池关键技术调研
- ◆ 编写技术方案和开发计划
- ◆ 开发基于 RDMA 的通讯层软件

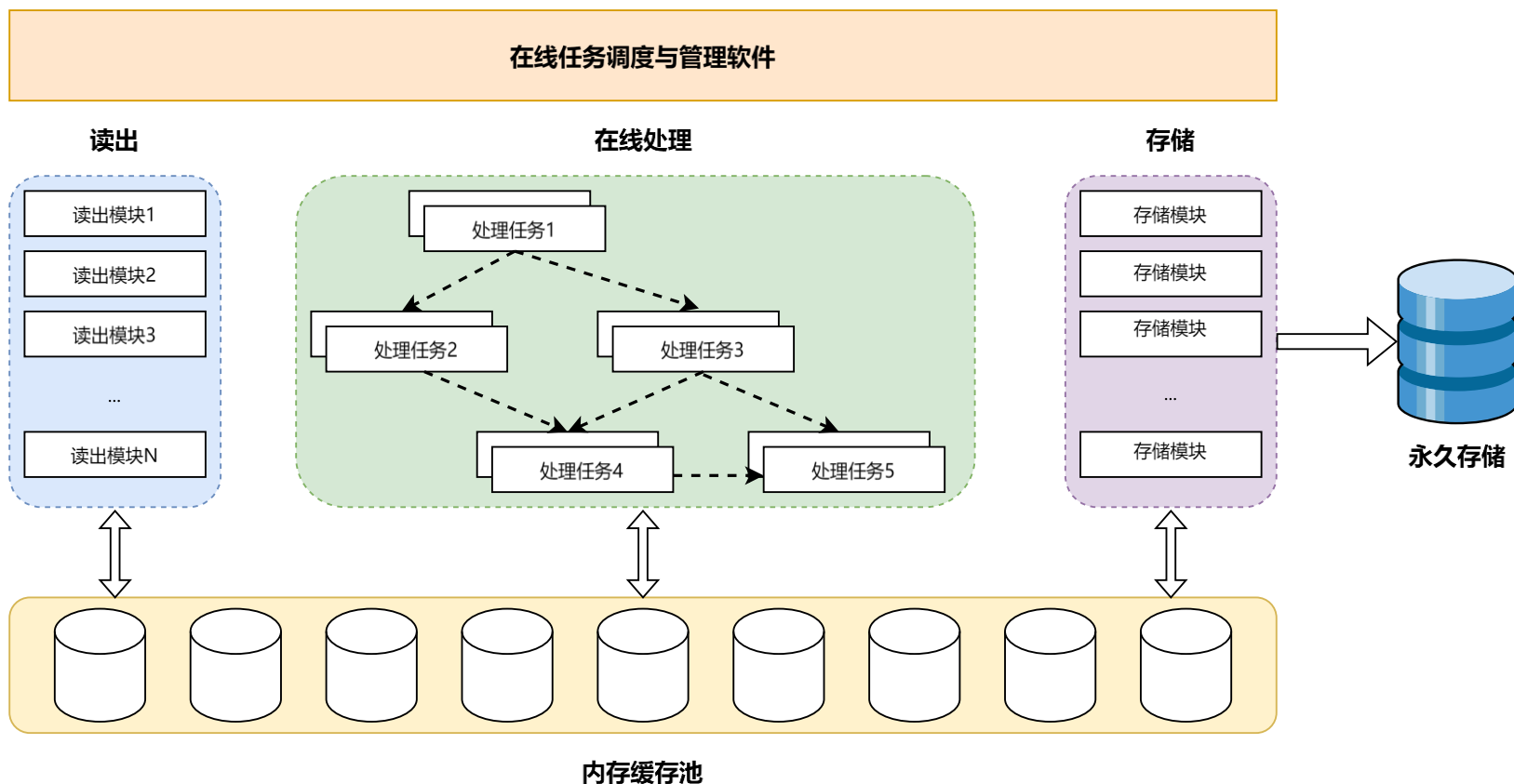
截止到2024.12

➤ 基于分布式内存缓存的在线计算方案

- ◆ 软件业务开发基本开发完毕



基于分布式内存缓存的在线计算方案



要求:

- 高性能
- 高效率
- 高可靠

关键技术:

- RDMA通讯层
- 分布式内存管理
- 分布式任务调度

开发要求:

- 基于 C++ 23 标准
- 尽量不使用外部依赖库
- 对文件操作的Trace



基于分布式内存缓存的在线计算方案

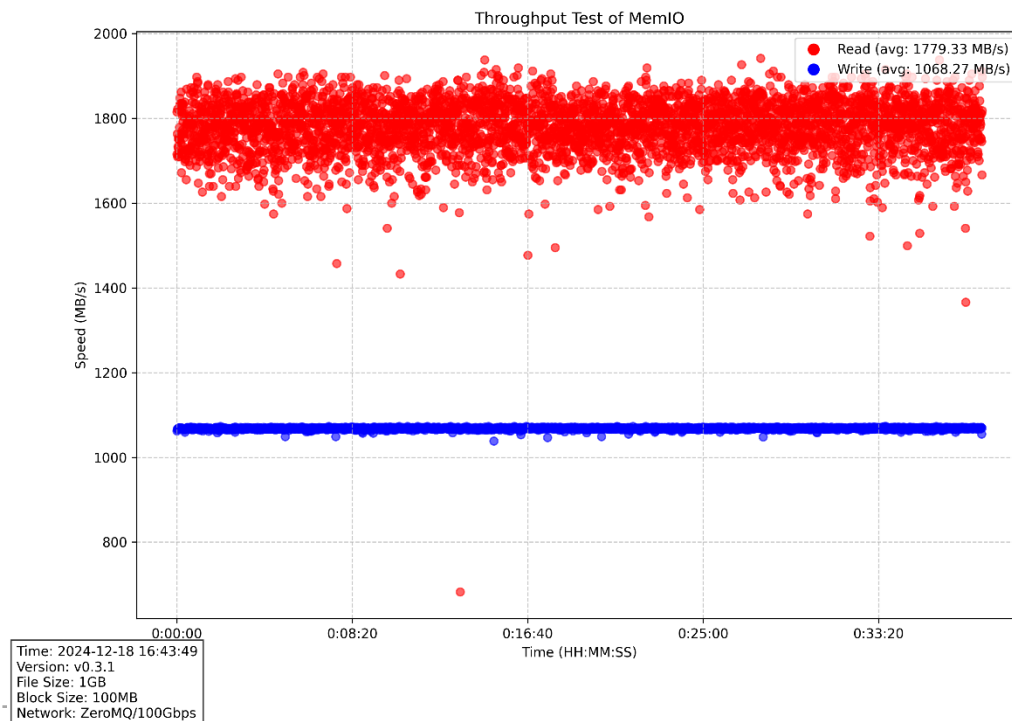
开发分布式内存缓存软件

- 2024.11.7 完成主要业务开发：v0.1版本
- 之后保持每两周迭代一个版本的开发计划

V 0.1	• 完成 主从节点通讯 • 完成 读写功能，性能500MB/s
V 0.2	• 优化 读写性能，性能800MB/s
V 0.3	• 优化 任务队列逻辑 • 添加 删除和列出功能 • 添加 CLI工具
V 0.4	• 添加 文件系统 • 添加 Mperf 测试工具 • 优化 TCP 读写性能

V 0.4 性能测试：

管理 4 个节点 800 GB 内存
100 Gbps TCP单线程，读 1.7GB，写 1GB
从节点 CPU ≤ 3 个核

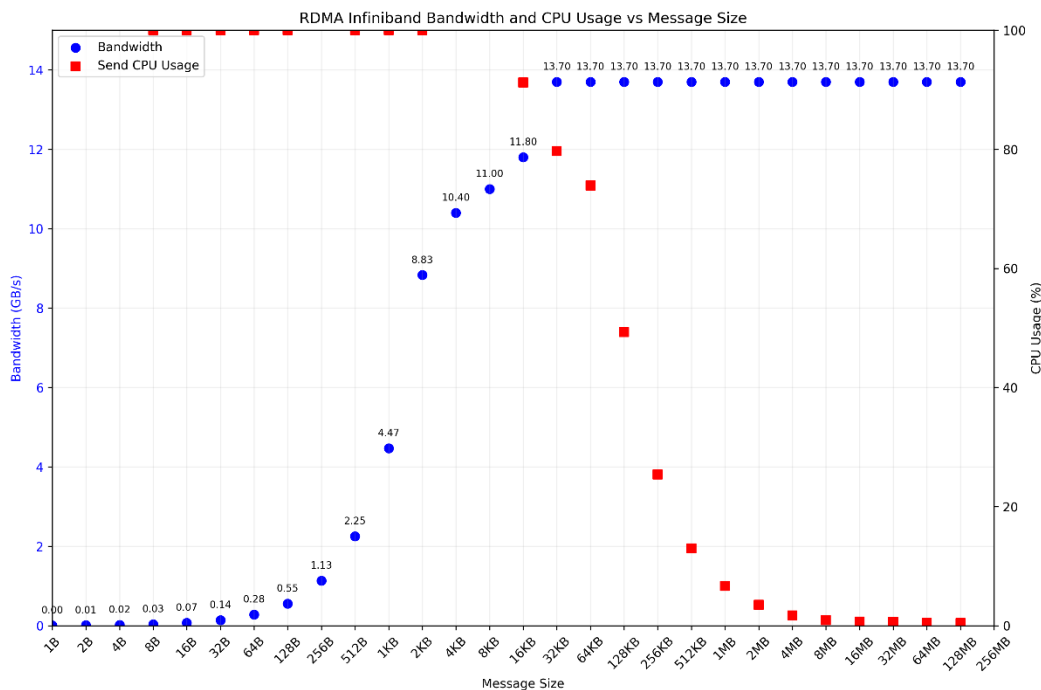




基于分布式内存缓存的在线计算方案

RDMA 研究

➤ InfiniBand 交换机、网卡安装与调试



➤ 基于 RoCE 的通讯层开发

```
ScopedMemory constructor:  
传输用时: 0.090984 秒  
传输速率: 10990.9 MB/s
```

测试环境: 100GbE

➤ 目前没有应用的原因:

- ◆ 需要 root 权限运行
- ◆ 需要 Linux 分页对齐
- ◆ 需要 使用 mmap 锁定内存防止被换出

对RDMA还需要进一步了解



总结与下一步计划

总结

➤ 内存缓存池:

◆ 11月份完成了第一个版本, 之后每两周迭代一个版本, 最新: V0.4, 读写 > 1 GB/s

➤ 合作组成员管理系统: 配合JUNO完成调试和测试工作

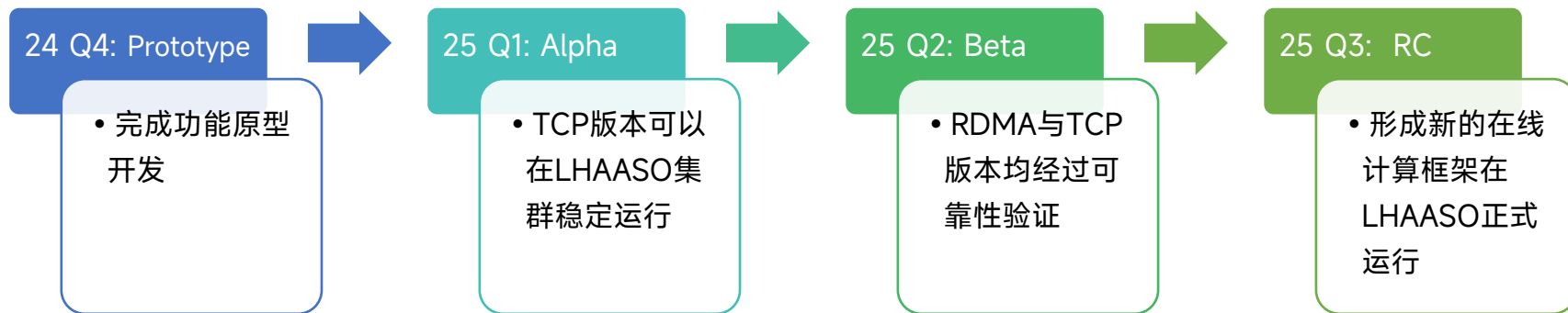
➤ IEEE TNS 论文已接收

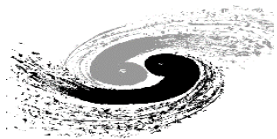
下一步计划

➤ TCP版本的缓存池软件在LHAASO上进行长期测试

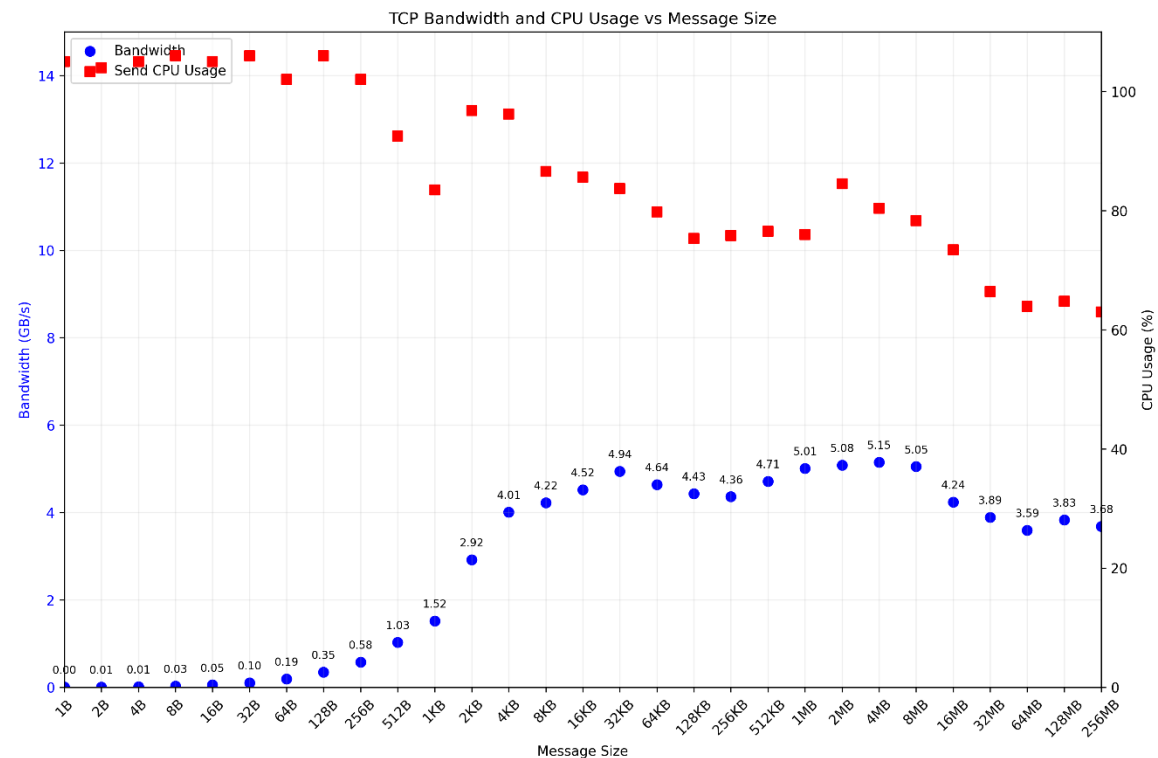
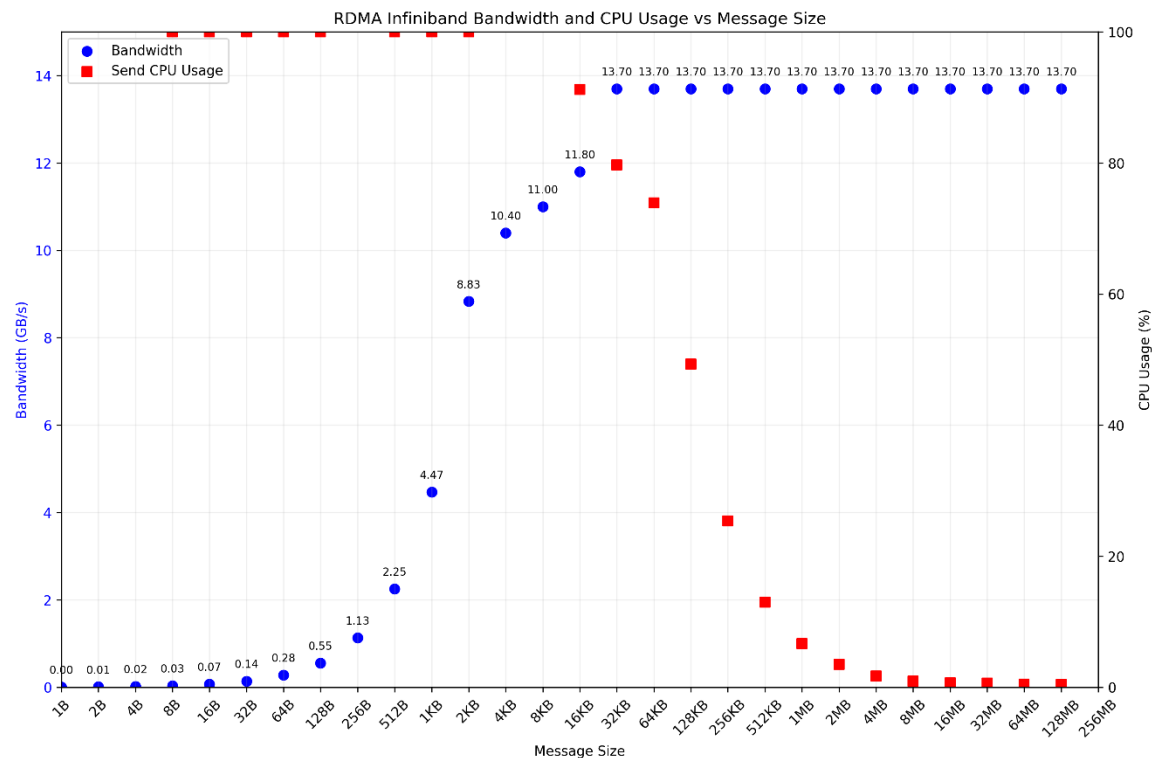
➤ RDMA版本的原型开发完毕

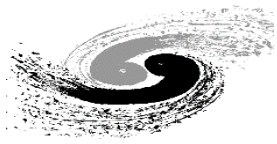
Hylia/TCP	2959.54 MB/s
Read	1750.24 MB/s
Write	1391.81 MB/s
Read (no hash)	2439.02 MB/s
Write (no hash)	1789.62 MB/s





backup





backup

