

PR #22420 完整报告

sgl-project/sclang

[Test] Add CPU unit tests for `MemoryPoolConfigurator`

合并时间: 2026-04-09 15:39

原文链接: <http://prhub.com.cn/sgl-project/sclang/pull/22420>

执行摘要

本 PR 为 MemoryPoolConfigurator 添加了 CPU-only 单元测试，通过模拟 ModelRunner 验证内存池配置逻辑，提升代码可靠性，无用户直接影响。

功能与动机

此 PR 基于 #22389 的后续工作，旨在添加测试以确保 MemoryPoolConfigurator 的内存使用率、页面对齐、比率正确性和约束行为得到验证，避免 GPU 依赖。

实现拆解

- 测试文件: test/registered/unit/model_executor/test_pool_configurator.py
- 关键函数:
 - _make_model_runner: 模拟 ModelRunner 对象
 - _full_per_token、_swa_per_token: 计算每 token 成本
 - _actual_memory_used: 计算实际内存使用
- 模拟技术: 使用 unittest.mock 模拟 GPU 相关函数，支持 CPU 环境测试。

评论区精华

review 评论为空，未发现讨论。

风险与影响

- 风险: 模拟环境可能不完全准确，测试覆盖可能不全面。
- 影响: 增加测试覆盖，对开发者有益，用户无感知。

关联脉络

- 直接关联 PR #22389，是其测试补充。
- 与其他内存或调度相关 PR (如 #22294) 可能有间接联系。