

PR #21533 完整报告

sgl-project/sclang

[AMD] Adjust AMD 4gpu partitions

合并时间: 2026-03-27 15:59

原文链接: <http://prhub.com.cn/sgl-project/sclang/pull/21533>

执行摘要

本次 PR 调整了 AMD 4GPU CI 测试的分区配置，将测试矩阵从两分区减至单分区，旨在优化 CI 执行效率。变更影响仅限于 CI 基础设施，风险较低，已于合并后关闭。

功能与动机

PR body 中未提供具体动机描述，仅包含贡献指南模板。基于标题和变更内容推断，可能目标是减少资源消耗或简化测试流程，例如缩短 CI 运行时间或修复分区问题。由于材料缺乏详细背景，具体原因不确定。

实现拆解

关键改动集中在两个 GitHub 工作流文件中：`-.github/workflows/pr-test-amd-rocm720.yml` 和 `.github/workflows/pr-test-amd.yml`：修改了 `jobs` 部分的矩阵配置。

- 将 `part: [0, 1]` 改为 `part: [0]`，意味着从并行运行两个分区减少到一个分区。
- 将 `--auto-partition-size 2` 改为 `--auto-partition-size 1`，调整了分区大小参数。这些变更一致应用于两个文件，属于 CI 配置的系统性调整，模块归类为 CI 基础设施。

评论区精华

Review 过程中仅有 reviewer [bingxche](#) 的批准，没有任何评论或讨论。这表明变更被快速接受，未引发技术争议或深度分析，团队可能认为这是例行维护。

风险与影响

风险分析：

- CI 配置变更可能减少测试覆盖，因为并行分区减半，可能导致某些测试用例未被完整执行。
- 如果分区调整不当，可能引发测试负载不均衡或超时问题，影响 CI 稳定性。
- 关键文件为两个工作流配置文件，错误修改可能中断整个 CI 流程。

影响分析：

- 对用户：无直接影响，用户不可见变更。
- 对系统：CI 测试的并行执行减少，可能增加单次测试时间，但总体资源消耗可能降低。

- 对团队：简化了 CI 配置，易于维护，但需监控测试结果以确保质量不受影响，影响程度为低。

关联脉络

与近期历史 PR 关联显示团队持续优化 CI 测试配置：

- PR #21516 ([CI] Fix nemotron nvfp4 test estimated time)：修复 CI 超时，同样调整测试参数。
- PR #21485 (Remove redundant DeepSeek V3 FP4 PCG test)：删除冗余测试以优化 CI 时间。这些 PR 共同反映了团队在 CI 基础设施上的改进趋势，旨在提升测试效率和资源利用率。本 PR 是该趋势的一部分，专注于 AMD 硬件相关的测试分区调整。