

PR #27352 完整报告

sgl-project/sglang

[AMD] fix(ci): run partition 3 of stage-c-test-large-8-gpu-amd

合并时间: 2026-06-05 15:06

原文链接: <http://prhub.com.cn/sgl-project/sglang/pull/27352>

执行摘要

修复 AMD CI 中 `stage-c-test-large-8-gpu-amd` 作业的矩阵配置, 将 `part` 从 `[0, 1, 2]` 扩展为 `[0, 1, 2, 3]`, 使分区 3 的测试用例不再被静默跳过。该 bug 源自 PR #24762 的配置遗漏。

功能与动机

`stage-c-test-large-8-gpu-amd` 作业使用 `--auto-partition-size 4` 将测试套房分割为 4 个分区, 但 CI 矩阵只调度了前 3 个分区 (`part: [0, 1, 2]`), 导致第 4 个分区 (约 1/4 的测试用例) 从未执行。被跳过的重要测试包括:

- `test_aiter_allreduce_fusion_amd.py`
- `test_aiter_allgather_amd.py`
- `test_deepseek_v3_mtp.py`
- `test_deepseek_v3_basic.py`

此问题由 PR #24762 (将 `--auto-partition-size` 从 3 改为 4) 引入, 但当时只更新了 `pr-test-amd-rocm720.yml` 的矩阵, 遗漏了 `pr-test-amd.yml` 的同步更新。

实现拆解

1. 修改 `.github/workflows/pr-test-amd.yml`: 将 `matrix.part` 的默认值从 `[0, 1, 2]` 改为 `[0, 1, 2, 3]`, 使得 4 个分区全部被 CI 调度。
2. 该修改仅涉及一行配置变更 (+1/-1), 但确保了测试覆盖的完整性。

YAML 配置关键行 (`pr-test-amd.yml`)

```
matrix:  
  part: [0, 1, 2, 3] # 原值为 [0, 1, 2], 缺少分区 3
```

评论区精华

无实质性 review 讨论。唯一评论来自 Gemini 自动回复, 表示无法审查文件类型。

风险与影响

- 风险: 极低。仅修正矩阵定义, 与源码逻辑无关。新增的分区 3 如包含不稳定测试可能偶尔导致 CI 失败, 但这正是 PR 希望暴露的真实问题。
- 影响: AMD CI 测试覆盖率恢复 100%, 关键测试不再被跳过。对用户无直接影响。

关联脉络

- PR #24762: 引入了 `--auto-partition-size 3 → 4` 的变更, 但未同步更新 `pr-test-amd.yml` 的矩阵, 为当前 bug 的根源。