

PR #24208 完整报告

sgl-project/sglang

[CI] Fix stage-b-test-4-gpu-b200 silently skipped, hanging wait-for-stage-b

合并时间: 2026-05-01 13:29

原文链接: <http://prhub.com.cn/sgl-project/sglang/pull/24208>

执行摘要

- 一句话: 修复 B200 CI 测试被静默跳过的问题
- 推荐动作: 建议快速合并。这是一个典型的 CI 回退不完整导致的 bug, 修复逻辑简单且正确。对于 CI 维护者, 这是个值得注意的教训: 回退时应检查所有引用点。

功能与动机

在 #24163 合并后, `stage-b-test-4-gpu-b200` 作业被静默跳过, 导致 B200 相关的 CI 测试完全缺失。`wait-for-stage-b` 因期望计数不匹配而超时, 阻塞了后续 CI 阶段。该问题已在实际 CI 运行中观察到 (如 job 73890630667 显示 `Progress: 26/26 jobs completed (expected 27)`, 最终超时)。

实现拆解

1. 在 `.github/workflows/pr-test.yml` 中, 将 `stage-b-test-4-gpu-b200` 作业的 `runs-on` 字段从 `${{ needs.check-changes.outputs.b200_low_disk_runner }}` 改回 `${{ needs.check-changes.outputs.b200_runner }}`。
2. 删除与之相关的三行注释, 这些注释描述了 `*-low-disk` 标签的作用, 因为该标签已不再被 `check-changes` 输出。
3. 无其他文件变更。

关键文件:

- `.github/workflows/pr-test.yml` (模块 CI 配置; 类别 `infra`; 类型 `infrastructure`): CI workflow 文件, 修改了 `stage-b-test-4-gpu-b200` 作业的 `runs-on` 配置, 修复了作业被静默跳过的问题。

关键符号: 未识别

评论区精华

无 review 讨论。但 PR body 详细记录了根因:

- 24163 回退时遗漏了消费端变更, 导致 `runs-on` 指向不存在的输出, GHA 静默跳过作业。
- `wait-for-stage-b` 因期望 27 个作业而永远等待。
- 暂无高价值评论线程

风险与影响

- 风险：风险极低：变更仅回退一行 runs-on 引用和删除三行注释，与 check-changes 的实际输出保持一致。不涉及任何业务逻辑或运行时变更。潜在风险是如果未来再次调整运行器标签，需要同步更新此处。
- 影响：直接影响：修复 CI 中 stage-b-test-4-gpu-b200 作业被跳过的问题，确保 B200 相关的 4-GPU 测试正常执行。间接影响：wait-for-stage-b 不再超时，后续 CI 阶段（如推理测试、端到端测试）能正常触发。仅影响 CI 流程，不涉及用户功能或系统性能。
- 风险标记：CI 修复，回退不完整

关联脉络

- PR #24163 Revert "[ci] split stage-c-test-4-gpu-b200 to enable a low-disk runner pool": 本 PR 修复了 #24163 回退时遗漏的消费端变更，是该回退的补全修复。
- PR #23505 [CI] Broaden stage-b-test-4-gpu-b200 runner pool to low-disk label: #23505 引入了 b200_low_disk_runner 输出，本 PR 回退了该 workflow 中的相应变更。