PR #23518 完整报告

sgl-project/sglang

test: move test_epd_disaggregation to nightly-4-gpu

合并时间：2026-04-23 11:59

原文链接：http://prhub.com.cn/sgl-project/sglang/pull/23518

执行摘要

- 一句话：将 EPD 解聚测试从提交门禁移至夜间套件，解决因精度边界抖动导致的 CI 阻塞。
- 推荐动作：该 PR 变更简单直接，主要价值在于 CI 流程优化。对于工程师，可快速浏览以了解测试套件调整模式；对于技术管理者，可关注其反映的测试抖动问题及后续修复计划。无需深入代码精读。

功能与动机

PR body 明确指出，test_mmmu 测试在精度边界存在抖动，断言 mmmu_accuracy > 0.40，但在生产 CI 中观测到恰好 0.40，导致 AssertionError: 0.4 not greater than 0.4 错误，阻塞了主分支的每提交 CI。为了在保留每日监控信号的同时解除阻塞，决定将整个测试文件从提交门禁套件移至夜间套件。

实现拆解

1. 变更测试注册配置：修改 test/registered/distributed/test_epd_disaggregation.py 文件中的 register_cuda_ci 调用，将 suite 参数从 "stage-c-test-4-gpu-h100" 改为 "nightly-4-gpu"，并添加 nightly=True 标志。
2. 保持硬件和测试逻辑不变：测试的估计时间（est_time=97）、硬件要求（4-GPU H100）以及所有测试类（如 TestEPDDisaggregationOmni）的实现均未改变，仅调整了执行频率。
3. 遵循现有模式：此变更模式与仓库中已有的夜间测试（如 test/registered/core/test_qwen3_next_deterministic.py）保持一致，确保 CI 注册解析无误。

关键文件：

- test/registered/distributed/test_epd_disaggregation.py（模块 分布式测试；类别 test；类型 test-coverage）：这是本次 PR 唯一变更的文件，通过修改 register_cuda_ci 调用，将 EPD 解聚测试从提交门禁套件移至夜间套件。

关键符号：未识别

关键源码片段

test/registered/distributed/test_epd_disaggregation.py

这是本次 PR 唯一变更的文件，通过修改 register_cuda_ci 调用，将 EPD 解聚测试从提交门禁套件移至夜间套件。

```
# 文件：test/registered/distributed/test_epd_disaggregation.py
# 关键变更：调整测试套件注册，从每提交门禁移至夜间运行，以解决抖动测试导致的 CI 阻塞。
# 原行（变更前）：
# register_cuda_ci(est_time=97, suite="stage-c-test-4-gpu-h100")
# 新行（变更后）：
register_cuda_ci(est_time=97, suite="nightly-4-gpu", nightly=True)
# 说明：
# - est_time=97：测试估计运行时间保持不变，确保资源分配一致。
# - suite="nightly-4-gpu"：将测试从提交门禁套件切换到夜间套件。
# - nightly=True：显式标记为夜间测试，遵循仓库现有模式（如 test_qwen3_next_deterministic.
py）。
# 影响：测试逻辑和硬件要求（4-GPU H100）未变，仅执行频率降低，解除 CI 阻塞。
```

评论区精华

本次 PR 的 review 讨论较为简单，仅有一条来自合并者 hnyls2002 的批准评论，内容为空，表明变更直接明了，无技术争议。PR body 中已详细说明了变更动机和测试计划，未引发额外讨论。

- 暂无高价值评论线程

风险与影响

- 风险：低风险：
  - 回归风险：测试逻辑本身未修改，仅调整执行套件，不会引入功能回归。
  - 信号丢失风险：由于测试移至夜间套件，对 test_mmmu 抖动的监控频率从每提交降至每日，可能延迟发现相关回归，但 PR body 已计划后续单独调查修复该抖动。
  - CI 配置风险：变更涉及 CI 套件配置，若 nightly-4-gpu 套件未正确设置或资源不足，可能导致测试无法运行，但 PR body 已验证本地解析并计划确认夜间运行。
  - 影响：
  - 对用户：无直接影响，此为内部测试调整。
  - 对系统：减少因抖动测试导致的 CI 失败，提高 PR 合并流程的稳定性。
  - 对团队：解除 CI 阻塞，加快开发迭代；但需关注夜间测试结果，及时跟进抖动修复。
  - 风险标记：测试信号延迟

关联脉络

- 暂无明显关联 PR