

PR #21897 完整报告

sgl-project/sglang

[CI] Increase multimodal server test timeout from 60 to 90 minutes

合并时间: 2026-04-02 11:19

原文链接: <http://prhub.com.cn/sgl-project/sglang/pull/21897>

执行摘要

该 PR 将 CI 工作流中多模态服务器测试的超时时间从 60 分钟延长至 90 分钟，以解决测试在输出结果时因超时中断的问题。变更仅涉及一个配置文件，风险较低，主要影响是提高 CI 测试的完整性和稳定性，但可能略微增加 CI 资源占用时间。

功能与动机

根据 PR 描述，在 1-GPU 分区中运行的 20 个扩散测试总耗时约 59 分钟，导致测试步骤在 60 分钟超时中断时正在打印 pytest 摘要结果，没有为结果输出留出时间余量。这影响了 CI 测试的完整执行和结果收集。延长超时旨在为测试运行和结果输出提供足够的时间缓冲，确保 CI 流程的可靠性。

实现拆解

该 PR 仅修改了 `.github/workflows/nightly-test-nvidia.yml` 文件，具体改动如下：

行号范围	变更内容	影响
510-510	<code>timeout-minutes: 60</code> → <code>timeout-minutes: 90</code>	将 1-GPU H100 配置的测试超时从 60 分钟延长至 90 分钟
568-568	<code>timeout-minutes: 60</code> → <code>timeout-minutes: 90</code>	将 2-GPU H100 配置的测试超时从 60 分钟延长至 90 分钟

这些修改直接增加了测试执行的最大允许时间，为测试运行和结果输出提供了更充裕的时间窗口。

评论区精华

该 PR 没有 review 评论或讨论，直接由作者合并。这表明变更简单直接，团队对调整超时以解决测试中断问题达成共识。

风险与影响

风险分析：

1. 可能掩盖性能回归：如果测试因性能下降而接近 90 分钟才完成，超时延长可能延迟发现性能问题。
2. CI 资源占用增加：测试运行时间可能延长，影响其他 CI 任务的调度效率。

影响分析：

1. 对 CI 的影响：确保多模态服务器测试能够完整执行并输出结果，提高 CI 可靠性。
2. 对团队的影响：减少因超时中断导致的虚假失败，提供更稳定的测试反馈。
3. 对用户和系统的影响：无直接影响，变更仅限于 CI 配置。

关联脉络

从近期历史 PR 看，该 PR 与以下 PR 相关：

- PR #21767：同样涉及多模态扩散模型 CI 测试的配置调整，关注测试覆盖和 CI 稳定性。
- PR #21898：同为 CI 测试配置调整，解决测试稳定性问题。
- PR #21882：涉及 CI 流程管理和策略，与本 PR 的 CI 配置调整共同构成基础设施维护的一部分。

这些 PR 共同反映了团队对 CI 测试稳定性和可靠性的持续关注，特别是在多模态和扩散模型测试领域。本 PR 的超时调整是这一系列基础设施优化中的一环，旨在解决测试执行时间与资源配置不匹配的问题。