

PR #29596 完整报告

sgl-project/sglang

Fix CI caused by <https://github.com/sgl-project/sglang/pull/29576>

合并时间: 2026-06-29 14:46

原文链接: <http://prhub.com.cn/sgl-project/sglang/pull/29596>

执行摘要

- 一句话: 修复 DSA indexer 测试因 device 参数错误导致的 CI 失败
- 推荐动作: 该 PR 是纯粹的 CI 修复, 简单直接。值得关注的是其揭示了测试代码应尽量贴近真实运行环境 (如 device 参数), 建议后续测试编写时注意此类细节。

功能与动机

29576 修复了 DSA indexer 的内存问题, 但其测试用例 `test_dsa_indexer.py` 中的 `_create_indexer` 方法在创建 `Indexer` 实例时没有指定 device, 导致 Indexer 在 CPU 上创建, CI 报错。本 PR 快速修复此问题, 使测试通过 (11 passed)。

实现拆解

在 `test/registered/kernels/test_dsa_indexer.py` 的 `_create_indexer` 方法中, 将 `indexer = Indexer(**params)` 包裹在 `with torch.device(self.device):` 上下文内, 确保 Indexer 直接在 CUDA 设备上初始化, 与后续 `indexer.to(device=self.device)` 保持一致。

关键文件:

- `test/registered/kernels/test_dsa_indexer.py` (模块 测试; 类别 test; 类型 test-coverage)
: 修复测试中 Indexer 创建时未指定 device 的问题, 通过 torch.device 上下文确保在 CUDA 上初始化。

关键符号: 未识别

评论区精华

无 review 讨论。PR author b8zhong 提交后, hnyls2002 触发 rerun-test, 测试通过后 PR 被合并。

- 暂无高价值评论线程

风险与影响

- 风险：变更仅影响测试代码，风险极低。但需注意，之前 PR #29576 的修复效果仍需通过该测试验证。
- 影响：仅影响 DSA indexer 测试的 CI 流程，用户无感知。确保持续集成通过，防止回归。
- 风险标记：测试配套变更

关联脉络

- PR #29576 Fix DSA indexer fusion bug causing excessive memory consumption.: 本 PR 修复了 #29576 引入的测试 CI 失败问题，是 #29576 的后续修复。