

PR #29671 完整报告

sgl-project/sglang

[AMD] Register fused_metadata_copy JIT kernel test for AMD nightly CI

合并时间: 2026-06-30 12:26

原文链接: <http://prhub.com.cn/sgl-project/sglang/pull/29671>

执行摘要

- 一句话: AMD CI 注册 fused_metadata_copy 测试
- 推荐动作: 该 PR 属于常规的 CI 覆盖补全, 技术复杂度低, 但体现了项目维护中对多平台兼容性的关注。值得关注的要点: JIT 内核的 ROCm 可移植性设计 (通过 `#ifndef USE_ROCM` 条件编译), 以及 CI 注册机制。

功能与动机

PR body 指出 fused_metadata_copy 内核测试目前仅运行在 NVIDIA 上, 但该内核是纯 int32 元数据拷贝, 且 JIT 源码已包含 ROCm 路径, 具备 ROCm 可移植性。PR 旨在填补 AMD nightly CI 中 JIT 内核测试的覆盖漏洞。

实现拆解

1. 导入调整: 在 test/registered/jit/test_fused_metadata_copy.py 中, 将 import 语句从 `from sglang.test.ci.ci_register import register_cuda_ci` 改为 `from sglang.test.ci.ci_register import register_amd_ci, register_cuda_ci`。
2. 注册添加: 在已有的两个 register_cuda_ci 调用之后, 新增一行 `register_amd_ci(est_time=100, suite="nightly-amd-kernel-1-gpu", nightly=True)`。该注册指向已有的 nightly-amd-kernel-1-gpu CI suite, 该 suite 已在 nightly-test-amd.yml 和 nightly-test-amd-rocm720.yml 中配置, 无需额外修改工作流。
3. 测试验证: PR 作者已在 MI3xx 硬件上通过 ROCm 7.0 和 ROCm 7.2 栈手动验证测试通过 (elapsed 49-52s, 远低于 est_time=100 超时阈值)。

关键文件:

- test/registered/jit/test_fused_metadata_copy.py (模块 测试; 类别 test; 类型 test-coverage; 符号 register_amd_ci): 唯一修改的文件, 添加了 AMD nightly CI 注册, 使 fused_metadata_copy 内核测试在 ROCm 平台上运行。

关键符号: 未识别

关键源码片段

`test/registered/jit/test_fused_metadata_copy.py`

唯一修改的文件，添加了 AMD nightly CI 注册，使 fused_metadata_copy 内核测试在 ROCm 平台上运行。

```
# 文件: test/registered/jit/test_fused_metadata_copy.py
# 变更: 添加 AMD nightly CI 注册, 使测试在 AMD ROCm 平台上运行

import sys
import time

import pytest
import torch

# 修改点: 增加 register_amd_ci 导入 (原仅导入 register_cuda_ci)
from sglang.test.ci.ci_register import register_amd_ci, register_cuda_ci

# NVIDIA 注册保持不变
register_cuda_ci(est_time=100, suite="base-b-kernel-unit-1-gpu-large")
register_cuda_ci(est_time=400, suite="nightly-kernel-1-gpu", nightly=True)

# 新增: AMD nightly 注册, suite 指向已有的 nightly-amd-kernel-1-gpu 流水线
# est_time=100 秒, 实际在 MI3xx 上实测耗时约 49-52 秒
register_amd_ci(est_time=100, suite="nightly-amd-kernel-1-gpu", nightly=True)
```

评论区精华

PR 无 review 评论, HaiShaw 直接批准。

- 暂无高价值评论线程

风险与影响

- 风险: 风险极低。变更仅添加一行 AMD CI 注册, 不影响任何运行时逻辑或 NVIDIA 注册。唯一的潜在风险是如果 Nightly AMD CI 环境缺少某些依赖 (如特定 HIP 版本), 测试可能失败, 但 PR 已在实际硬件上预先验证通过。
- 影响: 影响范围有限。仅对 AMD 平台 nightly 测试产生正向影响, 扩大了 fused_metadata_copy 内核的测试覆盖, 使其在 AMD 上也回归验证。NVIDIA 用户无影响。
- 风险标记: 暂无

关联脉络

- PR #29066 [CI] Migrate JIT tests to runner config registration: 本 PR 是 JIT 测试 CI 注册迁移工作的延续, 使用了相同的 register_amd_ci 机制。