

PR #30446 完整报告

sgl-project/sglang

[AMD] Register 2 CPU-bound 1-GPU tests (phase_checker, scripted_runtime_core) for AMD PR CI

合并时间: 2026-07-09 05:04

原文链接: <http://prhub.com.cn/sgl-project/sglang/pull/30446>

执行摘要

- 一句话: 为 AMD CI 注册两个 1-GPU 测试
- 推荐动作: 该 PR 属于基础设施层面的常规维护, 不涉及复杂逻辑, 可直接合并。值得关注的是其筛选测试的原则: 优先选取不依赖 CUDA 独占特性的测试, 并验证了实际运行结果, 为后续 AMD CI 扩展提供了参考模式。

功能与动机

增加 AMD PR CI 的测试覆盖范围, 确保 AMD GPU 上核心逻辑的正确性。PR body 说明两个测试已在 NVIDIA 上运行并已验证在 AMD 上通过。

实现拆解

1. 在 `test/registered/scripted_runtime/test_scripted_runtime_core.py` 中, 将 `register_cuda_ci` 的导入改为同时导入 `register_amd_ci`, 并在 `register_cuda_ci` 调用下方新增一行 `register_amd_ci(est_time=460, stage="stage-b", runner_config="1-gpu-small-amd")`。
2. 在 `test/registered/utils/test_phase_checker.py` 中进行完全相同的修改, `est_time` 设为 120。
3. 移除最初尝试包含的 `test_utils_update_weights.py`, 因为其 `setUpClass` 导入 `torch_memory_saver` 在 ROCm 上不可用, 导致 `ModuleNotFoundError`。

关键文件:

- `test/registered/scripted_runtime/test_scripted_runtime_core.py` (模块 运行时测试; 类别 test; 类型 test-coverage): 注册 AMD CI 运行入口, 新增一行 `register_amd_ci` 调用, 改动量小但影响 CI 覆盖。
- `test/registered/utils/test_phase_checker.py` (模块 阶段检查测试; 类别 test; 类型 test-coverage): 与上述文件相同模式, 为 `phase_checker` 测试添加 AMD CI 注册。

关键符号: 未识别

关键源码片段

`test/registered/scripted_runtime/test_scripted_runtime_core.py`

注册 AMD CI 运行入口，新增一行 register_amd_ci 调用，改动量小但影响 CI 覆盖。

文件: test/registered/scripted_runtime/test_scripted_runtime_core.py

```
import unittest

from sglang.srt.managers.schedule_batch import FINISH_ABORT
from sglang.test.ci.ci_register import register_amd_ci, register_cuda_ci # 新增 AMD 注册导入
from sglang.test.scripted_runtime.context import ScriptedContext
from sglang.test.scripted_runtime.http_server import ScriptedHttpServer
from sglang.test.scripted_runtime.req_handle import ScriptedReqHandle
from sglang.test.scripted_runtime.test_case import ScriptedTestCase
from sglang.test.scripted_runtime.chunked_helpers import (
    advance_to_decode_step,
    advance_to_nth_chunk,
    base_engine_kwargs,
    exhaust_row_pool,
    run_until_finished,
    warmup_radix,
)
from sglang.test.test_utils import CustomTestCase

# 注册 NVIDIA CI (原有)
register_cuda_ci(est_time=460, stage="base-b", runner_config="1-gpu-small")
# 注册 AMD CI (新增): 使用相同估算时间, 但指定 AMD 流水线阶段和 runner 配置
register_amd_ci(est_time=460, stage="stage-b", runner_config="1-gpu-small-amd")
```

test/registered/utils/test_phase_checker.py

与上述文件相同模式，为 phase_checker 测试添加 AMD CI 注册。

文件: test/registered/utils/test_phase_checker.py

```
import torch
from sglang.srt.utils.phase_checker import SimplePhaseChecker
from sglang.test.ci.ci_register import register_amd_ci, register_cuda_ci # 新增 AMD 注册导入
from sglang.test.test_utils import CustomTestCase

# 注册 NVIDIA CI (原有)
register_cuda_ci(est_time=120, stage="base-b", runner_config="1-gpu-small")
# 注册 AMD CI (新增)
register_amd_ci(est_time=120, stage="stage-b", runner_config="1-gpu-small-amd")
```

评论区精华

无 review 评论或讨论。

- 暂无高价值评论线程

风险与影响

- 风险：风险极低。两个测试均使用通用 PyTorch 操作或纯 CPU 逻辑，不涉及 CUDA 专用 API。导入的模块已在 NVIDIA CI 中验证，且 AMD CI 已在实际 ROCm 环境中通过了测试。排除的 `test_utils_update_weights.py` 避免了注册后因依赖缺失导致测试崩溃的风险。
- 影响：对用户无直接影响。对开发团队而言，AMD 硬件上的 CI 覆盖范围得到扩展，有助于更早发现回归问题。变更仅涉及测试注册，不影响任何生产代码。影响程度：小。
- 风险标记：暂无

关联脉络

- PR #30534 [AMD] Temporarily reduce AMD scheduled test frequency to save resource: 同一仓库近期对 AMD CI 的调整，反映了 AMD CI 资源管理策略的频繁变动。
- PR #30496 [AMD] Gate stage-c on stage-b-test-1-gpu-large-amd: 同样涉及 AMD CI 流水线门控，属于同一基础设施优化方向。