

PR #31371 完整报告

sgl-project/sglang

[CI] Remove nightly registrations redundant with scheduled stage runs

合并时间: 2026-07-16 05:12

原文链接: <http://prhub.com.cn/sgl-project/sglang/pull/31371>

执行摘要

- 一句话: 移除与定时调度冗余的 nightly CI 注册配置
- 推荐动作: 该 PR 已合并, 建议其他项目参考此清理思路。对于当前读者, 无需特殊关注, 但可学习其冗余分类的准则。

功能与动机

根据 PR body 说明: Tests registered in a PR stage already run in the 2x-daily scheduled full runs, 导致同一测试在三种场景下重复运行 (PR 阶段 + 定时全量 + nightly), 造成资源浪费。本 PR 旨在消除冗余, 同时确保必要的高参数覆盖不受影响。

实现拆解

1. 冗余识别: 分析 register_cuda_ci 调用, 发现 PR stage (如 stage="base-b-kernel-unit") 的测试已通过 run_all_tests=true 定时调度覆盖, 这些文件中的 nightly=True 注册为第三次重复。
2. 删除冗余注册: 对 32 个文件执行删除操作, 移除 nightly=True 的 register_cuda_ci 调用, 例如 test/registered/jit/test_tp_qknorm.py 和 test/registered/jit/diffusion/test_diffusion_modelopt_fp8_scaled_mm.py。
3. 保留必要覆盖: 对 19 个使用 get_ci_test_range 或依赖 SGLANG_JIT_KERNEL_RUN_FULL_TESTS 环境变量扩展参数范围的文件, 保留 nightly=True 注册, 并为部分文件添加注释说明原因 (如 test/registered/jit/test_dsv3_router_gemm.py)。AMD 和 NPU 相关的 nightly 注册保持不变。所有更改均在 test/registered/ 目录下的测试文件中。

关键文件:

- test/registered/jit/test_tp_qknorm.py (模块 CI 注册; 类别 test; 类型 test-configuration) : 示例删除了 nightly 注册, 是新规则的代表
- test/registered/jit/diffusion/test_diffusion_modelopt_fp8_scaled_mm.py (模块 CI 注册; 类别 test; 类型 test-configuration) : 示例删除了 nightly 注册, 属于扩散模型量化测试
- test/registered/jit/test_dsv3_router_gemm.py (模块 CI 注册; 类别 test; 类型 test-configuration) : 保留了 nightly 并添加注释说明, 是另一种情况
- test/registered/jit/deepseek_v4/test_c128_v2.py (模块 CI 注册; 类别 test; 类型 test-configuration) : 删除了 nightly, 但保留了 AMD nightly

- test/registered/jit/deepseek_v4/test_c4_v2.py (模块 CI 注册; 类别 test; 类型 test-configuration) : 与 test_c128_v2.py 类似
- test/registered/jit/deepseek_v4/test_fp4_indexer.py (模块 CI 注册; 类别 test; 类型 test-configuration) : 删除了 nightly 注册
- test/registered/jit/diffusion/test_fused_norm_scale_shift.py (模块 CI 注册; 类别 test; 类型 test-configuration) : 删除了 nightly 注册
- test/registered/jit/diffusion/test_group_norm_silu.py (模块 CI 注册; 类别 test; 类型 test-configuration) : 删除了 nightly 注册
- test/registered/jit/diffusion/test_qknorm_rope.py (模块 CI 注册; 类别 test; 类型 test-configuration) : 添加了 nightly 注册而非删除? 检查 patch 显示 +1/-0, 可能是添加注释? 实际是保留了 nightly 并添加注释
- test/registered/jit/diffusion/test_qwen_image_modulation.py (模块 CI 注册; 类别 test; 类型 test-configuration) : 添加了 nightly 注册? 类似
- test/registered/jit/diffusion/test_varlen_pack_pad.py (模块 CI 注册; 类别 test; 类型 test-configuration) : 添加了 nightly 注册?
- test/registered/jit/diffusion/test_varlen_uspattn_equivalence.py (模块 CI 注册; 类别 test; 类型 test-configuration) : 添加了 nightly 注册?

关键符号: 未识别

关键源码片段

test/registered/jit/test_tp_qknorm.py

示例删除了 nightly 注册, 是新规则的代表

```
# 变更后仅保留 base stage 注册, 删除了 nightly 注册
from sglang.test.ci.ci_register import register_cuda_ci

# 在 PR 测试和定时调度中运行一次
register_cuda_ci(
    est_time=300,
    stage="base-b-kernel-unit",
    runner_config="8-gpu-h200",
)
```

test/registered/jit/test_dsv3_router_gemm.py

保留了 nightly 并添加注释说明, 是另一种情况

```
# 保留 nightly 注册, 并通过注释说明必要性
from sglang.test.ci.ci_register import register_cuda_ci

register_cuda_ci(est_time=37, stage="base-b-kernel-unit", runner_config="1-gpu-large")
# Nightly is not redundant here: it sets SGLANG_JIT_KERNEL_RUN_FULL_TESTS=1
# to expand get_ci_test_range sweeps.
register_cuda_ci(est_time=148, suite="nightly-kernel-1-gpu", nightly=True)
```

评论区精华

该 PR 未产生 review 讨论（无人工评论）。作者在 issue 评论中只有一行机器人的配额提示。PR 本身意图清晰，无需争议。

- 暂无高价值评论线程

风险与影响

- 风险：主要风险：错误地删除必要覆盖可能导致回归未检测到。但作者通过筛选条件（仅删除完全重复的 nightlies；保留使用 `get_ci_test_range` 或 `SGLANG_JIT_KERNEL_RUN_FULL_TESTS` 的文件；保留 AMD/NPU 注册）降低了风险。此外，PR 在合并前已通过自身 CI，进一步验证了改动安全。潜在风险：未来添加新测试时若不了解此清理规则，可能错误添加冗余 nightly，但注释和代码规范可缓解。
- 影响：对用户：无直接影响。对开发者：CI 运行次数减少，排队时间缩短。对系统：节省 CI runner 资源（GPU 时间），尤其 8-GPU H200 等大型 runner。影响范围：中等，涉及 51 个测试文件，但均为 CI 配置变更，无功能影响。
- 风险标记：CI 冗余运行，测试覆盖减少

关联脉络

- 暂无明显关联 PR