

PR #28108 完整报告

sgl-project/sglang

[CI] Enforce modern `stage=`/`runner\_config=` form for dispatchable test suites

合并时间: 2026-06-13 08:32

原文链接: <http://prhub.com.cn/sgl-project/sglang/pull/28108>

执行摘要

- 一句话: 为 CUDA 测试强制现代 stage/runner_config 注册形式, 支持 /rerun-test 调度
- 推荐动作: 值得精读, 尤其是 scripts/ci/check_registered_tests.py 中的 AST 解析和正则校验设计, 可作为同类预提交检查的参考。

功能与动机

PR body 指出: 以 {stage}-test-{runner_config} 命名的套件如果通过旧式 suite= 单字符串注册, 无法被 /rerun-test 调度 (调度器只解析现代 runner_config= 形式及 CPU、nightly-/weekly- 旧形式)。

实现拆解

1. 检测旧式 suite= 并阻止: 在 scripts/ci/check_registered_tests.py 中, 解析 register_cuda_ci 调用, 若发现 suite 参数与正则 `^(.+)-test-(.+)$` 匹配且未同时设置 stage/runner_config, 则报错并提示改用现代形式。
2. 迁移 3 个测试注册: 将 test/registered/tokenizer/test_multi_detokenizer.py、test/registered/unit/spec/test_eagle_worker_v2_topk1_fastpath.py 和 test/registered/vlm/test_token_id_retokenize_e2e.py 中的 register_cuda_ci(est_time=..., suite="base-b-test-...") 替换为 register_cuda_ci(est_time=..., stage="base-b", runner_config="1-gpu-..."), 功能等价, 但使这些测试变得可调度。
3. 更新 pre-commit 钩子名称: 在 .pre-commit-config.yaml 中将该钩子的名称从 check registered tests have CI registry 改为 validate registered test CI registries, 更准确地反映其新增职责。

关键文件:

- scripts/ci/check_registered_tests.py (模块 CI 脚本; 类别 infra; 类型 infrastructure; 符号 main, _MODERN_SHAPE) : 核心文件: 添加了 pre-commit 守卫, 通过 AST 解析检测旧式 suite= 注册, 并强制 CUDA 测试使用现代形式。
- test/registered/tokenizer/test_multi_detokenizer.py (模块 测试; 类别 test; 类型 test-coverage) : 1 行变更: 将 register_cuda_ci(est_time=211, suite="base-b-test-1-gpu-large") 替换为现代形式。

- test/registered/unit/spec/test_eagle_worker_v2_topk1_fastpath.py (模块 测试; 类别 test; 类型 test-coverage) : 1 行变更: 将 suite="base-b-test-1-gpu-small" 改为 stage="base-b", runner_config="1-gpu-small"。
- test/registered/vlm/test_token_id_retokenize_e2e.py (模块 测试; 类别 test; 类型 test-coverage) : 1 行变更: 将 suite="base-b-test-1-gpu-large" 改为 stage="base-b", runner_config="1-gpu-large"。
- .pre-commit-config.yaml (模块 配置; 类别 config; 类型 configuration) : 1 行变更: 更新 pre-commit 钩子名称描述, 与新增职责对齐。

关键符号: main

关键源码片段

scripts/ci/check_registered_tests.py

核心文件: 添加了 pre-commit 守卫, 通过 AST 解析检测旧式 **suite=** 注册, 并强制 CUDA 测试使用现代形式。

```
import re
from scripts.ci.ci_register import ut_parse_one_file, HWBackend

# 匹配 {stage}-test-{runner_config} 形状的套件名
_MODERN_SHAPE = re.compile(r"^(.+)-test-(.+)$")

def main() -> int:
    # ... 省略相同部分 ...
    legacy_shape = []
    for f in files:
        registries, _ = ci_register.ut_parse_one_file(f)
        if not registries:
            continue
        for r in registries:
            # 只关心 CUDA + 旧式 suite 且未用 stage/runner_config
            if not (r.backend == HWBackend.CUDA and r.suite is not None
                    and r.stage is None and r.runner_config is None):
                continue
            m = _MODERN_SHAPE.match(r.suite)
            if m:
                legacy_shape.append((f, r.suite, m.group(1), m.group(2)))
    # ... 输出错误信息并退出码 ...
```

评论区精华

无 review 评论, 核心讨论体现在 PR body 和提交信息中: 作者明确指出现代形式与旧形式在调度上的差异, 并设计了 pre-commit 守卫来保证一致性。决定仅对 CUDA 测试且套件名匹配 **{stage}-test-{runner_config}** 模式时强制现代形式, 其他情况保留旧形式。

- 现代 stage/runner_config 形式 vs 旧式 suite 形式对 /rerun-test 调度的影响 (design): 将 3 个 CUDA 测试迁移到现代形式, 并添加 pre-commit guard 防止未来引入类似问题。

风险与影响

- 风险：

1. pre-commit 误报：若未来有合法的旧式 suite= 命名恰好符合 {stage}-test-{runner_config} 模式但无法使用现代形式（例如特殊字符），会被误拦截。但当前设计允许 AMD/CPU/NPU 和 nightly 套件继续使用旧式，减少了误报面。
2. 迁移测试的行为无变化：三个迁移的测试仅更改注册参数，CIRegistry.effective_suite 会将两个形式解析为相同套件字符串，因此调度和执行行为完全不变。- 影响：对开发者：在 CI 相关测试注册时需使用现代形式，pre-commit 会提前拦截违规。对系统：使 base-b 套件下的 CUDA 测试（如 test_eagle_worker_v2_topk1_fastpath.py）可被 /rerun-test 单独重跑，提高了调试效率。影响范围：仅涉及 CI 注册逻辑，不改变产品代码。

- 风险标记：pre-commit 误报可能，迁移测试行为等价

关联脉络

- 暂无明显关联 PR