

PR #3148 完整报告

radixark/miles

[CI] temporarily disable failing FT tests

合并时间: 2026-09-09 07:30

原文链接: <http://prhub.com.cn/radixark/miles/pull/3148>

执行摘要

- 一句话: 临时禁用 5 个失败的容错端到端测试, 保持 CI nightly 绿灯。
- 推荐动作: 这是一个低风险的 CI 维护操作, 用于管理已知测试失败。它值得精读, 但主要是为了确认禁用的测试范围正确, 并理解“未来 FT 交付”的跟踪机制。对于普通工程师, 了解此变更即可, 无需深入代码细节。

功能与动机

根据 PR body, "Nightly run 34243026014 on main 3de96596f failed 5 ft-short tests in stage-c-8-gpu-h200. Mark them disabled="will enable in future FT delivery" so nightly stays green until the next FT delivery re-enables them." 这表明失败的测试是已知问题, 需要在未来的 FT 交付中修复或重新启用。

实现拆解

1. 变更入口: 在 5 个端到端测试入口文件 (如 tests/e2e/ft/test_trainer_ft_deterministic_dp2_cp2_pp2.py 等) 中调用 register_cuda_ci 的地方。
2. 核心变更: 在每个文件的 register_cuda_ci(...) 调用中, 新增 disabled="will enable in future FT delivery" 关键字参数。此参数由 CI 注册机制读取, 用于标记测试应被跳过。
3. 变更范围: 仅涉及 tests/e2e/ft/ 目录下的 5 个特定测试文件, 每个文件仅新增 1 行。没有修改任何源码、配置或部署脚本。
4. 配套说明: 这是一个临时性变更, 不涉及测试逻辑本身。test_qwen36 (4-gpu 作业) 未被触及。

关键文件:

- tests/e2e/ft/test_trainer_ft_deterministic_dp2_cp2_pp2.py (模块 确定性容错测试; 类别 test; 类型 test-coverage) : 被禁用的 5 个端到端测试之一, 展示了如何通过 disabled 参数跳过测试。
- tests/e2e/ft/test_trainer_ft_with_failure_dp2_cp2_tp2_ep2.py (模块 故障容错测试; 类别 test; 类型 test-coverage) : 被禁用的 5 个端到端测试之一, 属于故障场景测试。

关键符号: 未识别

关键源码片段

tests/e2e/ft/test_trainer_ft_deterministic_dp2_cp2_pp2.py

被禁用的 5 个端到端测试之一，展示了如何通过 `disabled` 参数跳过测试。

```
from tests.ci.ci_register import register_cuda_ci
from tests.e2e.ft.conftest_ft.scenario_deterministic import run_ci

# 注册 CI 测试并标记为禁用
register_cuda_ci(
    est_time=2900,
    suite="stage-c-8-gpu-h200",
    labels=["ft-short"],
    disabled="will enable in future FT delivery", # 禁用原因说明
)

_MODE: str = "dp2_cp2_pp2"
```

评论区精华

本次 PR 的 review 讨论非常简短。主要评论包括：

- claude[bot]: 提供了标准的代码审查提示。
- fzyzcjy: 表示 "LGTM given it is just disabling as discussed offline", 表明这是一个线下讨论后达成的共识操作。
- Zhichenzzz: 空评论表示同意。 核心结论：这是一个预先讨论过的、简单的 CI 维护操作，没有技术争议。
- 暂无高价值评论线程

风险与影响

- 风险：
 1. 测试覆盖风险：禁用这 5 个测试会暂时减少容错功能的端到端测试覆盖，直到它们被重新启用。
 2. 技术债务风险：如果“未来的 FT 交付”被延迟或遗忘，这些测试可能会长期处于禁用状态，掩盖潜在的回归问题。
 3. 回归风险：此变更本身不引入新功能或修改现有逻辑，因此直接回归风险极低。 - 影响：用户影响：无直接影响。此变更仅影响 CI 流水线行为。系统影响：CI nightly 构建的稳定性得到改善，避免因已知失败测试导致的红灯。团队影响：减少了 CI 失败告警的噪音，让团队能专注于其他问题。但需要跟踪后续的 FT 交付以重新启用这些测试。
- 风险标记：测试覆盖缺失，技术债务

关联脉络

- 暂无明显关联 PR