

# PR #43188 完整报告

vllm-project/vllm

[ci] Revert model executor test back to L4

合并时间: 2026-05-20 14:01

原文链接: <http://prhub.com.cn/vllm-project/vllm/pull/43188>

## 执行摘要

- 一句话: 将 model executor 测试从 H200 回退到 L4
- 推荐动作: 此 PR 为临时性 CI 配置调整, 关注度低。建议在驱动更新后恢复 H200 测试, 或采纳 bot 建议显式指定 L4 以增加确定性。

## 功能与动机

Nvidia driver needs update for H200 MIG, 因此将 model executor 测试从 H200 迁移回 L4, 以保证 CI 正常运行。

## 实现拆解

1. 在 `.buildkite/test_areas/model_executor.yaml` 中删除 `device: h200_35gb` 一行, 使测试不再指定设备, 默认由 buildkite 分配到 L4 agent。
2. 代码变更仅为一行删除, 无其他文件改动。

关键文件:

- `.buildkite/test_areas/model_executor.yaml` (模块 CI 配置; 类别 config; 类型 configuration) : 删除了 `device: h200_35gb` 配置行, 是本次变更的唯一文件, 直接决定了测试的硬件分配。

关键符号: 未识别

## 评论区精华

Gemini code-assist bot 评论指出, PR 标题是回退到 L4, 但删除 `device` 行可能导致测试被分配到非 L4 agent, 建议显式设置 `device: l4` 以确保正确路由。该建议未被采纳。

- 显式指定 L4 设备 (design): 当前 PR 未采纳建议, 直接删除了 `device` 行。后续可能因路由不确定性导致问题。

## 风险与影响

- 风险: 删除 `device` 后, 测试可能被分配到其他类型的 agent (如 A100), 导致测试环境不一致或资源占用冲突。如果 L4 资源不足, 可能阻塞 CI 流程。

- 影响：影响 CI 基础设施：Model Executor 测试将从 H200 设备迁移到默认设备（预期为 L4），测试性能可能因 GPU 不同而发生变化，但功能不受影响。
- 风险标记：缺少显式设备指定，临时性配置

## 关联脉络

- PR #43129 [ci] Move language models tests (hybrid) back to L4: 同样是将测试从 H200 迁移回 L4，属于同系列 CI 调整。