

# PR #24325 完整报告

sgl-project/sglang

rerun-test: route deepep h200 suite to deepep runner

合并时间: 2026-05-04 06:57

原文链接: <http://prhub.com.cn/sgl-project/sglang/pull/24325>

## 执行摘要

本 PR 将 rerun-test 中 `stage-c-test-deepep-8-gpu-h200` 阶段的硬件映射从 `8-gpu-h200` 改为 `8-gpu-h200-deepep`，并新增对应 runner 配置，确保 rerun 请求落在正确的专用 runner 上。同时添加了一个临时的 `FORCE_REBUILD_DEEPEP` 环境变量以兼容 Torch 升级。

## 功能与动机

PR body 明确说明: Match the runner that pr-test.yml uses for `stage-c-test-deepep-8-gpu-h200` (`8-gpu-h200-deepep`)，以便 `/rerun-test` 命令在该套件上使用与原始 PR job 相同的硬件。此前映射错误导致 rerun 路由到普通 H200 runner 而非 deepep 专用 runner。

## 实现拆解

- `.github/workflows/rerun-test.yml`: 在 runner 列表的 `8-gpu-h200` 之后新增 `8-gpu-h200-deepep` 条目。同时在 `env` 中添加 `FORCE_REBUILD_DEEPEP: '1'`，并标注为临时措施，用于当前 Torch 升级 PR 的兼容性，合并到 main 前需回退。
- `scripts/ci/utis/slash_command_handler.py`: 在 `handle_rerun_stage` 函数的映射字典中，将 `stage-c-test-deepep-8-gpu-h200` 对应的值从 `8-gpu-h200` 改为 `8-gpu-h200-deepep`。这样当用户输入 `/rerun-test stage-c-test-deepep-8-gpu-h200` 时，命令会路由到正确的 runner 标识。

## 评论区精华

该 PR 无人工 review 评论，只有 bot 的自动确认，表明改动简单且无歧义。

## 风险与影响

- 风险: 极低。仅影响 CI 的 rerun 路由逻辑，不涉及任何业务代码。唯一的潜在风险是 `FORCE_REBUILD_DEEPEP` 临时环境变量若忘记移除，可能会在后续 CI 中引入不必要的重建开销。
- 影响: 确保 deepep 测试 rerun 的硬件一致性，避免因 runner 不同导致的测试失败或结果偏差。

## 关联脉络

与同仓库近期 PR 对比，本 PR 属于典型的 CI 基础设施维护。类似 PR 如 #24265（删除 `--prerelease allow`）和 #24282（修复夜间 CI 并发组）都体现了团队对 CI 稳定性的持续投

入。