

# PR #7186 完整报告

PaddlePaddle/FastDeploy

[CI] Use GPU-Build-RL runner for \_build\_linux\_rl.yml

合并时间: 2026-04-03 20:24

原文链接: <http://prhub.com.cn/PaddlePaddle/FastDeploy/pull/7186>

## 执行摘要

此 PR 将 RL 构建工作流的 runner 从 GPU-Build 改为 GPU-Build-RL，并更新基础镜像和 Paddle whl URL，以解决资源不匹配问题。它属于 CI 基础设施优化，影响范围限于构建环境，但需关注依赖兼容性和 review 中发现的配置错误修复情况。

## 功能与动机

为什么做: 根据 PR body，当前 runner 配置与 RL 构建任务的 GPU 需求不匹配，可能导致资源错配或执行问题。具体表述为: "Current runner configuration does not match the GPU requirements of RL build tasks, which may lead to resource mismatch or execution issues." 因此，需要调整 runner 以确保正确的资源分配。

## 实现拆解

实现涉及两个关键文件的修改:

### 1. [.github/workflows/\\_build\\_linux\\_rl.yml](#):

- 将 runs-on 从 [self-hosted, GPU-Build] 改为 [self-hosted, GPU-Build-RL]。
- 更新基础镜像版本，从 paddlecloud-ubuntu24.04-gcc13.3-cuda12.9-cudnn9.9-bccl1.4.1.4-nccl2.26.5-openmpi4.1.5-FleetY13.0.0-rc2 改为 paddlecloud-ubuntu24.04-gcc13.3-cuda12.9-cudnn9.9-bccl1.4.1.4-nccl2.26.5-openmpi4.1.5-FleetY13.0.0-v2.4.0-rc1。
- 更新 PADDLE\_WHL\_URL 以指向新的构建流水线。

### 2. [.github/workflows/ce\\_job.yml](#):

- 同步更新 PADDLE\_WHL\_URL，确保一致性。

## 评论区精华

review 讨论中，fastdeploy-bot 指出 PR 描述与实际变更不符，并报告了关键问题:

🔗 Bug分支名拼写错误: `devel` 应为 `develop` 此更改会导致 HPU CI 无法在 `develop` 分支的 PR 上触发，CI 流程将完全失效。

🔗 Bug注释掉了主要 CI 测试流程 此变更注释掉了以下关键 CI 步骤: ... 这些是核心 CI 保护措施，合入后将导致 PR 无法通过常规 CI 检查。

这些评论强调了 CI 配置变更的严谨性，最终在提交前被修正（基于文件列表仅显示两个文件变更）。

## 风险与影响

风险：

- 资源分配风险：如果 GPU-Build-RL runner 配置不当，可能影响 RL 构建性能。
- 依赖更新风险：基础镜像和 Paddle whl URL 变更可能引入兼容性问题。
- CI 流程风险：review 中提到的错误如果未修复，将导致 CI 失效。

影响：

- 对用户无直接影响。
- 对系统：优化 RL 构建资源分配，可能提升构建效率；但需监控构建稳定性。
- 对团队：需要适应新的 CI 配置，影响程度中等。

## 关联脉络

与近期 PR 的关联：

- PR #7138 和 #7132：同属 CI 基础设施优化，涉及 Docker 配置和测试环境改进，可对比学习。
- PR #7171：涉及 RL 模块的 bug 修复和优化，本 PR 的 CI 变更可能支持此类任务的构建。

整体上，此 PR 反映了团队对 CI 环境持续优化的趋势，特别是在资源分配和依赖管理方面。