

# PR #41668 完整报告

vllm-project/vllm

[Build] Switch CUDA 12.9 wheel builds to PyTorch manylinux\_2\_28 base

合并时间: 2026-05-16 04:46

原文链接: <http://prhub.com.cn/vllm-project/vllm/pull/41668>

## 执行摘要

- 一句话: CUDA 12.9 车轮构建切换到 manylinux\_2\_28 基础镜像
- 推荐动作: 该 PR 属于基础设施维护, 逻辑清晰、改动极小, 无需精读。但可作为 CI/CD 镜像策略变更的参考。

## 功能与动机

作为 #41416 的后续跟进, 将 cu130 wheel 构建的 manylinux\_2\_28 切换镜像到 cu129, 以确保 cu129 发布的 wheel 也使用 glibc 2.28 基础镜像, 匹配 PyTorch 官方 wheel 的 ABI 兼容性要求。PR body 明确指出“mirror the cu130 wheel build switch onto cu129”。

## 实现拆解

在 .buildkite/release-pipeline.yaml 中对 CUDA 12.9 的两个 wheel 构建步骤 (ARM64 和 x86\_64) 分别添加了 `--build-arg BUILD_OS=manylinux` 和 `--build-arg BUILD_BASE_IMAGE=...` 两个参数, 将基础镜像从默认版本切换为 PyTorch 的 manylinux\_2\_28 builder 镜像。ARM64 步骤使用 `pytorch/manylinuxaarch64-builder:cuda12.9`, x86\_64 步骤使用 `pytorch/manylinux2_28-builder:cuda12.9`。其他配置 (如 CUDA 架构列表、构建命令等) 保持不变。

关键文件:

- .buildkite/release-pipeline.yaml (模块 CI 配置; 类别 config; 类型 configuration) : 唯一的变更文件, 修改了 CUDA 12.9 两个 wheel 构建步骤的 Docker 构建参数

关键符号: 未识别

## 评论区精华

无实质 review 讨论, 仅有自动化机器人评论。

- 暂无高价值评论线程

## 风险与影响

- 风险: 风险极低: 仅修改两个构建步骤的基础镜像参数, 不涉及任何源代码或运行时逻辑。若镜像不兼容或构建失败, 只会影响 cu129 wheel 的发布, 不会影响已有 wheel 或运行时容器。

- 影响：影响范围限于 cu129 wheel 的构建产物，用户将获得具有 glibc 2.28 最低版本要求的 wheel，提升了跨 Linux 发行版的兼容性。对现有用户无直接影响，仅影响后续版本发布。
- 风险标记：影响范围小

## 关联脉络

- PR #41416 [Build] Switch CUDA 13.0 wheel builds to PyTorch manylinux\_2\_28 base:  
本 PR 是 #41416 的镜像切换，将 cu130 的 manylinux\_2\_28 切换策略应用到 cu129。