

PR #38413 完整报告

vllm-project/vllm

[ROCm] [Release] Update ROCm variant from rocm700 to rocm721

合并时间: 2026-03-28 14:07

原文链接: <http://prhub.com.cn/vllm-project/vllm/pull/38413>

执行摘要

本 PR 更新 ROCm 变体至 7.2.1, 调整 CI/CD 发布流水线配置, 确保版本一致性, 是 PR #38252 的后续工作, 风险较低, 主要涉及基础设施维护。

功能与动机

跟随 PR #38252, 将 ROCm 版本从 7.0.0 升级到 7.2.1, 以保持发布流程的最新性和兼容性。PR body 中明确提到此 PR 是 #38252 的后续, 目的是更新 ROCm 变体。

实现拆解

主要修改两个文件:

- `.buildkite/release-pipeline.yaml`: 将 VARIANT 从 `rocm700` 改为 `rocm721`, 并添加 if 条件 `if: build.env("NIGHTLY") == "1"` 限制 nightly 镜像发布。
- `tools/vllm-rocm/generate-rocm-wheels-root-index.sh`: 更新默认 VARIANT 值从 `rocm700` 到 `rocm721`。这些变更确保 CI/CD 流水线使用正确的 ROCm 变体, 避免配置不一致。

评论区精华

review 中, `gemini-code-assist[bot]` 指出一个关键问题:

"This inconsistency is critical. It will cause wheels built with ROCm 7.0 to be published under a path that indicates they are for ROCm 7.2.1. This will likely lead to incorrect dependencies and runtime failures for users." 作者 `tjtanaa` 回复: "This is the case before I synced to main. I have rebased, now it is pointing to rocm 7.2.1" 讨论焦点是版本一致性的正确性风险, 结论是问题在同步 main 分支后已解决。

风险与影响

风险: 如果版本不一致未修复, 可能导致发布错误版本的 wheel, 用户使用时出现依赖错误或运行时失败。但作者确认已修复, 风险较低。影响:

- 对系统: 更新 CI/CD 发布流程, 确保使用 ROCm 7.2.1 版本。
- 对用户: 修复后无直接影响; 未修复可能导致兼容性问题。影响范围限于 ROCm 相关构建和发布, 不涉及核心模型逻辑。

关联脉络

此 PR 与 PR #38252 直接关联，后者更新了 Docker 基础镜像至 ROCm 7.2.1。结合近期历史 PR 如 #38367（文档更新）和 #37453（ROCm bugfix），可见 vllm 仓库正持续优化 ROCm 支持，体现了基础设施版本管理的演进趋势。