

PR #55354 完整报告

vllm-project/vllm

[ROCm][CI] Bump ROCk release image build timeout to 3h

合并时间: 2026-09-05 00:16

原文链接: <http://prhub.com.cn/vllm-project/vllm/pull/55354>

执行摘要

- 一句话: 将 ROCk release 镜像构建超时时间从 60 分钟增加至 180 分钟, 以解决冷缓存构建导致的 CI 超时问题。
- 推荐动作: 这是一个低风险、高必要性的 CI 配置修复。对于关心 CI 稳定性和 ROCm 支持的工程师, 建议快速浏览以了解问题背景和解决方案。对于关注 vLLM 发布流程或 AMD ROCm 集成的管理者, 此 PR 证明了团队对 CI 基础设施问题的快速响应能力。

功能与动机

直接动机是修复 nightly 发布流程阻塞问题。根据 PR body 描述, 在引入 ROCm 10.0 基础镜像 (PR#55246) 后的首次 nightly 运行 (release-v2 #6097) 中, 'Build release image - x86_64 - ROCK' 任务在编译 vLLM 的 csrc 和 NIXL/mori/lmcache/deepep 组件时, 因使用全新的 TheRock 工具链缓存 (cold sccache) 而耗时超过 60 分钟限制, 最终超时。这阻塞了 **nightly-rocm100** 的发布。PR 中提到, 同一构建在缓存预热后的后续运行中可在 60 分钟内完成, 因此需要一个更长的超时来容纳首次构建。

实现拆解

1. 变更入口与范围: 变更仅涉及 CI 配置文件 .buildkite/release-pipeline.yaml 中的一个参数, 范围极小且明确。
2. 核心逻辑调整: 在文件的第 1037 行, 将 build-rock-release-image 任务的 timeout_in_minutes 从硬编码的 60 修改为 180。此任务是 ROCk release Docker 镜像构建流程中的第二个步骤, 负责在拉取的 ROCm 10.0 基础镜像上编译并构建最终的 vLLM 发布镜像。
3. 配套与验证: 变更没有代码逻辑修改, 因此无需新增单元测试。验证方式是依赖后续的 nightly 运行来检验超时延长后任务是否能成功完成, 正如 PR body 所述: “Effect verified by the next nightly release-v2 run。”同时, 作者提供了简单的 YAML 语法检查命令作为本地测试。

关键文件:

- .buildkite/release-pipeline.yaml (模块 构建流水线; 类别 config; 类型 configuration): 这是本次 PR 唯一修改的文件, 包含了 ROCk release 镜像构建任务的完整配置, 其中关键的超时参数在此处被调整。

关键符号: 未识别

评论区精华

PR 的讨论内容较少且无技术争议。主要评论来自机器人自动总结和 CI 触发指令。审阅者 [dllehr-amd](#) 以 “Sounds good” 简单批准，表明该变更被视为一个直接且必要的配置调整，无需深入讨论。社区机器人 [coderabbitai\[bot\]](#) 也未提出任何具体改进建议。

- 暂无高价值评论线程

风险与影响

- 风险：

1. CI 资源占用风险：超时时间提升至 180 分钟，意味着即使构建失败，该任务也可能占用 CI 代理长达 3 小时，相比之前的 1 小时有所增加。然而，PR body 解释了这是为缓存预热留出必要空间的合理权衡，且 “warm runs finish well under it”，因此风险可控。
2. 配置错误风险：这是一个单行配置修改，直接风险极低。主要的潜在风险是数值设置不当（过短或过长），但 180 分钟是基于实际观察的编译时间（60.8 分钟仍不足）和工程经验设定的保守值。

- 影响：

1. 对系统的影响：直接影响是稳定了 ROCm 10.0 平台的 nightly 构建和发布流程。移除了因构建超时导致的发布阻塞，确保了 ROCm 10.0 用户能够持续获得更新的 vLLM Docker 镜像。
 2. 对用户的影响：间接影响。对于使用 ROCm 10.0 的 vLLM 用户，这保证了他们所依赖的 Docker 镜像能够按期发布，避免了因 CI 问题导致的镜像缺失或延迟。
 3. 对团队的影响：减少了维护团队处理因 CI 超时导致的 nightly 构建失败和手动干预的需求。同时，这也是对 ROCm 10.0 基础镜像迁移（PR#55246）的配套修复，使整个迁移更加完整和可靠。
- 风险标记：CI 任务超时调整

关联脉络

- PR #55317 [CI/Build] Fix flaky failures in CPU CI image building: 同为修复 CI/CD 流程中的瞬态或配置问题，以确保构建稳定性。两者都属于近期 CI 基础设施维护的一部分。
- PR #55136 [CI] Raise AMD Spec Decode Eagle 1 job timeout to 35min: 同样是因为特定任务运行时间超出预设限制而调整超时参数的 CI 配置变更，反映了仓库根据实际运行情况持续调优 CI 流水线的趋势。