

# PR #43136 完整报告

vllm-project/vllm

[ROCm] Bump ROCm to 7.2.3

合并时间: 2026-05-29 00:42

原文链接: <http://prhub.com.cn/vllm-project/vllm/pull/43136>

## 执行摘要

- 一句话: ROCm 7.2.3 升级, 移除 profiler hotfix
- 推荐动作: 值得快速审阅并通过, 该 PR 是常规的平台依赖升级, 逻辑清晰、改动集中, 且经过充分验证。

## 功能与动机

ROCm 7.2.3 版本已包含 profiler hotfix, 因此不再需要为 7.2.2 重建 CLR, 简化 Dockerfile 并减少构建耗时。

## 实现拆解

1. 升级基础镜像: 在 docker/Dockerfile.rocm\_base 中将 BASE\_IMAGE 从 rocm/dev-ubuntu-22.04:7.2.2-complete 改为 7.2.3-complete。
2. 移除 CLR 热补丁: 删除 Dockerfile 中用于 7.2.2 profiler hotfix 的全部 RUN 指令 (克隆 rocm-systems、编译 CLR、安装等约 20 行), 因为该修复已集成到 7.2.3。
3. 更新 Triton/PyTorch 分支: 将 TRITON\_BRANCH 和 PYTORCH\_BRANCH 更新到对应 ROCm 7.2.3 的兼容提交 (分别更改为 0f380657 和 96bfee12)。
4. 修复 aiter 构建: 在 Dockerfile 中将 git clone --recursive 与 git checkout \${AITER\_BRANCH} 合并为单条带有 --branch 的 clone 命令。
5. 更新发布流水线: 在 .buildkite/release-pipeline.yaml 中将环境变量 VARIANT 从 rocm722 改为 rocm723, 确保 wheel 索引正确标识版本。

关键文件:

- docker/Dockerfile.rocm\_base (模块 构建脚本; 类别 infra; 类型 infrastructure): 核心变更文件: 升级基础镜像版本, 移除旧版本 profiler 热补丁, 更新 Triton/PyTorch 分支, 修复 aiter 构建步骤。
- .buildkite/release-pipeline.yaml (模块 部署配置; 类别 config; 类型 configuration): 发布流水线配置文件, 将环境变量 VARIANT 从 rocm722 改为 rocm723, 确保 wheel 索引正确。

关键符号: 未识别

## 关键源码片段

## docker/Dockerfile.rocm\_base

核心变更文件：升级基础镜像版本，移除旧版本 profiler 热补丁，更新 Triton/PyTorch 分支，修复 aiter 构建步骤。

```
# 从 rocm/dev-ubuntu-22.04:7.2.2-complete 升级到 7.2.3-complete
ARG BASE_IMAGE=rocm/dev-ubuntu-22.04:7.2.3-complete
# 更新 Triton 到与 ROCm 7.2.3 兼容的提交
ARG TRITON_BRANCH="0f380657"
# 更新 PyTorch 到 release/2.11 分支 (5/13 版本)
ARG PYTORCH_BRANCH="96bfee12"

# 移除以下所有行：旧版本 profiler hotfix 的 CLR 重建步骤
# RUN apt-get update && apt-get install -y rocm-llvm-dev
# RUN pip install CppHeaderParser
# RUN git clone --no-checkout ... && ... && make install && rm -rf /tmp/rocm-systems

# ... 中间部分不变 ...

# 修复 aiter 构建：将分开的 clone 和 checkout 合并为单条带 branch 的 clone
RUN git clone --recursive --branch ${AITER_BRANCH} ${AITER_REPO}
RUN cd aiter \
    && git submodule update --init --recursive \
    && pip install -r requirements.txt
```

## 评论区精华

gemini-code-assist[bot] 提出 PyTorch 2.11 分支与 TorchVision v0.24.1 可能存在 API 不兼容风险，但该疑虑未得到人工 reviewer 回应或确认。dllehr-amd 要求等待 @micah-wil 确认不再需要 profiler hotfix，micah-wil 随后通过回归测试验证了修复效果，dllehr-amd 因此批准。tjtanaa 则直接要求将 **VARIANT** 改为 **rocm723**，并在修改后批准。

- PyTorch 与 TorchVision 版本兼容性 (correctness): 未得到人工回应，但 PR 最终被批准，表明风险在可接受范围内。
- 确认不需 profiler hotfix (question): micah-wil 通过回归测试证实 7.2.3 通过 profiler 测试，无需 hotfix，dllehr-amd 随后批准。
- VARIANT 环境变量值 (design): micah-wil 按要求修改，tjtanaa 随后批准。

## 风险与影响

- 风险：主要风险在于 PyTorch 2.11 与 TorchVision v0.24.1 的兼容性，若出现构建失败或运行时错误，可能影响 ROCm 用户的推理流程。但由于该 PR 专注于基础设施升级且经作者测试通过，实际风险较低。
- 影响：影响范围限于 ROCm 平台用户。移除 CLR 构建步骤可缩短 Docker 镜像构建时间，提升 CI/CD 效率。升级后所有 ROCm 7.2.x 用户将受益于内置 profiler 修复。
- 风险标记：依赖版本兼容性

## 关联脉络

- 暂无明显关联 PR