

# PR #24981 完整报告

sgl-project/sglang

[AMD] avoid aiter re-installing triton in amd\_install\_dependency

合并时间: 2026-05-11 23:05

原文链接: <http://prhub.com.cn/sgl-project/sglang/pull/24981>

## 执行摘要

- 一句话: 修复 AMD CI 中 aiter 重建时 triton 被重装问题
- 推荐动作: 值得合入, 但可考虑 gemini-code-assist[bot] 的建议进一步优化构建隔离设置, 以提升 CI 效率。

## 功能与动机

避免 aiter 在 AMD CI 重建时重新安装 triton, 因为 aiter 的 setup.py 在 commit #3086 后无条件调用 install\_triton.sh, 导致卸载 CI 镜像自带的 ROCm 优化 triton 并重新安装, 可能引发兼容性问题并掩盖更重要的 bug。

## 实现拆解

1. 设置环境变量: 在 scripts/ci/amd/amd\_ci\_install\_dependency.sh 中 aiter 重建命令前添加 AITER\_USE\_SYSTEM\_TRITON=1, 使得 aiter 的 setup.py 跳过 triton 安装流程。
2. 改用 pip install -e .: 将 python3 setup.py develop 替换为 pip install -e ., 以匹配已有 Dockerfile 和 aiter 自身 CI workflow, 利用 PEP 517 构建隔离避免依赖冲突。
3. 验证: CI 运行 [run 25676429617] 显示成功保留系统 triton 3.4.0。

关键文件:

- scripts/ci/amd/amd\_ci\_install\_dependency.sh (模块 CI 脚本; 类别 infra; 类型 infrastructure) : 唯一变更文件, 通过新增环境变量和 CLI 命令变更避免 triton 重装

关键符号: 未识别

## 关键源码片段

`scripts/ci/amd/amd_ci_install_dependency.sh`

唯一变更文件, 通过新增环境变量和 CLI 命令变更避免 triton 重装

```
# sglang/scripts/ci/amd/amd_ci_install_dependency.sh # 上下文: 当 NEED_REBUILD 为 true 时执行 aiter 重建 if [{"${NEED_REBUILD}"=="true"}];then # ... 其他准备步骤
dockerexecci_sglangbash-c" cd /sgl-workspace/aiter && \ # 设置环境变量避免 aiter 重新安装 triton (保留 CI 镜像自带的 ROCm 优化版) AITER_USE_SYSTEM_TRITON=1
GPU_ARCHS=${GPU_ARCH_LIST}python3 setup.py develop " echo"[CI-AITER-CHECK]
=== AITER REBUILD COMPLETE ===" fi 注意: 实际 diff 中该命令可能为 pip install -e .,
```

但核心思路一致。

## 评论区精华

gemini-code-assist[bot] 指出使用默认构建隔离的 `pip install` 在 CI 中效率低下，会重新下载超过 1GB 的 `torch` 等依赖，建议增加 `--no-build-isolation` 并使用缓存和 `install_with_retry` 提高效率。该建议未被采纳，但当前方案已经通过 `AITER_USE_SYSTEM_TRITON=1` 规避了核心问题。

- 暂无高价值评论线程

## 风险与影响

- 风险：风险极低：仅修改一行环境变量和安装命令，不影响业务代码。潜在风险是如果未来 `aiter` 的安装流程变化，环境变量名可能失效，但 CI 日志可快速发现。
- 影响：直接影响 AMD CI 的 `aiter` 构建步骤，缩短构建时间（避免下载 `triton`），提高 CI 稳定性。不影响用户功能或运行时行为。
- 风险标记：暂无

## 关联脉络

- PR #24924 [AMD] Pin cache-dit==1.3.0 in rocm.Dockerfile + AMD CI install script: 同为 AMD CI 依赖修复，同一脚本文件变更