

PR #22657 完整报告

sgl-project/sclang

[AMD] Remove aiter hotfixes in Dockerfile covered by aiter v0.1.12.post1

合并时间: 2026-04-13 15:01

原文链接: <http://prhub.com.cn/sgl-project/sclang/pull/22657>

执行摘要

该 PR 移除了 AMD ROCm Dockerfile 中针对 aiter v0.1.10.post3 的两个临时补丁，因为这些补丁已由上游 aiter v0.1.12.post1 版本修复。变更简化了 Dockerfile 维护，风险较低，仅影响 AMD 平台的构建环境。

功能与动机

PR body 明确指出，aiter 已升级至 v0.1.12.post1，其中包含了上游修复（如 PR#2173、PR#2169、PR#2575），解决了之前需要在 `docker/rocm.Dockerfile` 中手动打补丁的问题。因此，移除这些冗余的临时解决方案，以减少技术债务并提升代码清晰度。

实现拆解

仅修改一个文件：`docker/rocm.Dockerfile`。删除了以下补丁：

- 针对 ROCm 7.2 的 if False: 补丁：原用于禁用 `_gluon_deepgemm_fp8_paged_mqa_logits` 函数（因布局错误），现已由 aiter PR#2575 修复。
- 针对所有 GPU 架构的 `dict(config)` 补丁：原用于解决 `get_gemm_config()` 键缺失错误，现已由 aiter PR#2173 和 PR#2169 修复。

变更后，Dockerfile 更简洁，依赖关系更清晰。

评论区精华

Review 中仅有一名审核者（HaiShaw）批准，无评论讨论。这表明变更基于明确的上游修复，团队一致认可其必要性。

风险与影响

- 风险：依赖上游 aiter 修复的正确性；若修复不完整，可能影响 AMD GPU 功能。移除补丁后，Dockerfile 构建对 aiter 版本更敏感，需确保后续升级兼容性。
- 影响：仅影响 AMD ROCm 平台的 Docker 构建环境，简化维护，可能轻微提升构建速度。对用户无直接影响。

关联脉络

与近期 PR 如 #21908 (Intel GPU 依赖升级) 和 #21441 (CUDA 版本升级) 类似, 均涉及基础设施的依赖管理和清理。这反映了团队持续优化构建配置、减少临时补丁的趋势。关联 Issue (#2169、#2173、#2575) 展示了上游 aiter 的修复过程, 为本 PR 提供了直接依据。