

PR #36892 完整报告

sgl-project/sglang

[AMD] release rocm10 image for gfx1250 from amd_helios

合并时间: 2026-08-29 03:00

原文链接: <http://prhub.com.cn/sgl-project/sglang/pull/36892>

执行摘要

- 一句话: 新增 gfx1250 每日发布作业, 从 amd_helios 分支构建镜像
- 推荐动作: 该 PR 为基础设施变更, 代码量小, 逻辑清晰。可快速浏览, 重点了解新增作业的独立性和版本标记策略。

功能与动机

PR body 指出 main 分支的 `docker/rocm.Dockerfile` 尚无 gfx1250 阶段, 需要从 `amd_helios` 分支构建 gfx1250 镜像, 以支持 MI45x 的提前集成。同时, 将作业独立于 `publish`, 避免 bring-up 构建失败阻塞正式镜像发布。

实现拆解

1. 新增触发选项: 在 `release-docker-amd-rocm10.yml` 的 `workflow_dispatch` 中增加 `publish-gfx1250` 选项。
2. 新增作业定义: `publish-gfx1250` 作业, 使用 `amd-docker-scale runner`, 环境 `prod`, 条件判断 `job_select` 为 `all` 或 `publish-gfx1250`。
3. 构建步骤: `checkout amd_helios` 分支 (`fetch-depth: 0`) , 设置 Python 3.10, 计算日期和版本, 确定 `pretend_version`, 登录 Docker Hub, 执行 `docker build` 并使用 `--build-arg` 指定 `SGL_BRANCH`、`BUILD_TYPE=all`、`GPU_ARCH=gfx1250-rocm1000` 等参数。
4. 推送镜像: 将构建的镜像打标签为 `v<version>-rocm10-mi45x-dev-<date>` 并推送到 `rocm/sgl-dev`。

关键文件:

- `.github/workflows/release-docker-amd-rocm10.yml` (模块 发布配置; 类别 `infra`; 类型 `infrastructure`) : 新增 gfx1250 每日发布作业, 定义构建流程和触发条件。

关键符号: 未识别

评论区精华

无 review 讨论。作者在 issue 评论中提供了已发布的 action 运行链接。

- 暂无高价值评论线程

风险与影响

- 风险：风险较低：
 - 构建从 amd_helios 分支拉取代码，若该分支不稳定，可能影响镜像质量，但镜像标记为 dev，不会作为正式发布。
 - 新增作业独立运行，不会阻塞现有正式发布流程。
 - 该作业仅在 amd-docker-scale runner 上运行，需确保该 runner 池可用。
 - 影响：影响范围：AMD GPU 用户，特别是 MI45x 使用者，可获取每日构建的开发镜像用于测试。CI 基础设施增加了每日构建任务，但独立于现有发布。团队可提前验证 gfx1250 支持。
 - 风险标记：依赖 bring-up 分支，不影响正式发布

关联脉络

- PR #36434 [AMD] Add ROCm 10 (gfx942 / gfx950) release images: 同 workflow 文件，本 PR 在其基础上新增 gfx1250 作业。
- PR #36775 CI: split JIT kernel unit tests into two partitions: 均为基础设施调整，虽细节不同但属于 CI 优化范畴。