

PR #27765 完整报告

sgl-project/sglang

[AMD-miles] add amd-miles daily docker build workflow

合并时间: 2026-06-16 08:22

原文链接: <http://prhub.com.cn/sgl-project/sglang/pull/27765>

执行摘要

- 一句话: 为 AMD Miles 项目添加 nightly Docker 构建 workflow
- 推荐动作: 建议合并。该 PR 解决了 Miles 项目的基础设施需求, 且方案简单、复用现有配置。合并后可考虑添加失败通知 (如 Slack) 并定期检查 secret 有效性。其他团队如需类似 nightly 构建可参考此模式。

功能与动机

Miles 项目需要 nightly ROCm 镜像以支持持续集成。由于 sglang 仓库已配置好 AMD Docker 构建 runner 和 rocm/sgl-dev 推送凭证, 因此将 Miles 的 nightly 镜像构建 workflow 直接添加到 sglang 仓库, 避免重复设置 CI 基础设施。引用 PR body: 'Miles needs a nightly ROCm image. Since the AMD Docker build runners and rocm/sgl-dev push credentials are already set up in sglang, this PR adds the Miles nightly image build workflows here.'

实现拆解

1. 新增两个 GitHub Actions workflow 文件: release-docker-amd-miles-rocm700-nightly.yml (MI300/gfx942, ROCm 7.0) 和 release-docker-amd-miles-rocm720-nightly.yml (MI355X/gfx950, ROCm 7.2)。
2. 每个 workflow 包含 publish 和 push_local_registry 两个 job。publish job 检出 radixark/miles@main, 通过 docker/Dockerfile.rocm 构建镜像, 并推送 dated tag (如 miles-rocm720-mi35x-20250301) 到 rocm/sgl-dev。
3. push_local_registry job 在 GPU runner 上拉取该镜像并推送至内部 registry (10.245.143.50:5000), 供 AMD CI 使用。
4. 通过 --build-arg MILES_COMMIT 传递精确的 miles commit SHA, 确保 Dockerfile 内 clone 版本与外部 checkout 一致。
5. workflow 通过 schedule 定时 (每天 UTC 12:00) 触发, 也支持 workflow_dispatch 手动触发。并发组设置避免重叠运行。

关键文件:

- .github/workflows/release-docker-amd-miles-rocm700-nightly.yml (模块 CI workflow; 类别 infra; 类型 infrastructure): 定义 Miles 在 ROCm 7.0 (gfx942) 环境下的 nightly 镜像构建流程, 包括从 miles 仓库检出、构建、推送和镜像复制。

- `.github/workflows/release-docker-amd-miles-rocm720-nightly.yml` (模块 CI 工作流; 类别 infra; 类型 infrastructure) : 与 `rocm700` 类似, 但针对 ROCm 7.2 (gfx950) 环境。

关键符号: 未识别

评论区精华

评论中主要讨论了两点:

- commit SHA 一致性: bingxche 发现外部 checkout 记录的 SHA 与 Dockerfile 内 clone 可能不一致, 建议通过 `--build-arg MILES_COMMIT` 传递精确值。XinyuJiangCMU 采纳并修复。
- Tag 策略: bingxche 询问是否预期同时推送 rolling tag, XinyuJiangCMU 认为只保留 dated tag 更好, 已移除 rolling tag。
 - 确保 Miles 内部 clone 与外部 checkout 使用相同的 commit SHA (correctness): XinyuJiangCMU 确认修复, 通过新增步骤记录 SHA 并作为 build arg 传入。
 - rolling tag vs dated tag 策略 (design): 移除 rolling tag, 只保留 dated tag, 与仓库惯例对齐。

风险与影响

- 风险:
 - 依赖外部仓库: 构建依赖 `radixark/miles@main`, 若 miles 仓库发生 breaking change 或不可达, 工作流会失败。建议添加失败告警。
 - 凭证管理: 工作流使用 `DOCKERHUB_AMD_USERNAME` 和 `DOCKERHUB_AMD_TOKEN`, 需确保这些 secret 安全且定期轮换。
 - 内部网络依赖: `push_local_registry` job 依赖内部 registry 地址 `10.245.143.50:5000`, 仅适用于 AMD 内部网络, 外部 contributor 无法验证。
 - 镜像存储消耗: 构建后镜像默认保留 1 天, 但构建过程可能消耗较多 runner 磁盘空间。
- 影响:
 - 用户影响: 无, 仅影响 Miles 项目和 AMD CI。
 - 系统影响: 每天两次 Docker 构建和推送, 增加 CI runner 负载和存储消耗 (artifact 保留 1 天)。
 - 团队影响: AMD 团队获得自动化镜像构建, 减少手动操作。工作流托管在 sglang 仓库, 可能增加维护责任。
 - 风险标记: 外部依赖, 凭证安全, 内部网络依赖

关联脉络

- PR #27839 Fixes lint on branches: 本 PR 合并前依赖了 #27839 修复 lint 问题, 以确保 CI 通过。