

PR #26731 完整报告

sgl-project/sclang

[NPU] Update documentation for software version upgrades

合并时间: 2026-06-06 20:06

原文链接: <http://prhub.com.cn/sgl-project/sclang/pull/26731>

执行摘要

本 PR 将 NPU (Ascend) 软件栈版本从 CANN 8.5.0 升级到 9.0.0, 同步更新相关依赖版本号, 并修改文档、Dockerfile 和 CI workflows 以反映这些变化。属于常规版本升级配套更新, 无新增功能或修复。

功能与动机

随着 Ascend 驱动栈和库的更新, SGLang 需要支持最新的 CANN 9.0.0、PyTorch Adapter 26.0.0 等版本。本 PR 基于 #26466 的变更, 更新了安装文档、快速入门指南、最佳实践文档中的版本要求和命令示例, 同时调整 Docker 构建参数和 CI 默认镜像, 确保用户和 CI 使用一致的新版依赖。

实现拆解

- 系统要求文档 (ascend_npu.mdx) : 更新 HDK、CANN、PyTorch Adapter、MemFabric、Triton、SGLang NPU Kernel 和 MemFabric-zbal 的版本号, 并替换对应下载链接。
- 快速入门文档 (ascend_npu_quick_start.mdx) : 将每日构建镜像标签从 main-cann8.5.0-a3 改为 main-cann9.0.0-a3, 稳定版标签同步更新。
- 最佳实践文档 (ascend_npu_best_practice.mdx) : 删除两处 export PATH=/usr/local/Ascend/8.5.0/compiler/bishengir/bin:\$PATH, 因新版不再需要。
- Docker 构建配置 (npu.Dockerfile) : 将默认 CANN_VERSION 参数从 8.5.0 改为 9.0.0。
- CI 工作流 (full-test-npu.yml、nightly-test-npu.yml) : 将使用的 Docker 镜像默认标签从 cann:8.5.0-a3 改为 cann:9.0.0-a3。

评论区精华

reviewer ping1jing2 直接 approve, 未产生讨论线程。变更清晰, 无需额外解释。

风险与影响

- 风险: CANN 主版本升级 (8.5.0 → 9.0.0) 可能引入未知兼容性问题, 但该 PR 仅更新文档和 CI, 实际兼容性验证需在后续测试中确认。文档和配置均为文本替换, 不涉及运行时变更。

- 影响：用户需按照新版文档安装依赖；CI 将使用新镜像运行测试。若镜像仓库未及时部署新版本，可能导致 CI 失败。

关联脉络

本 PR 是 #26466 的后续，用于同步文档和 CI 配置，确保版本升级后的流程一致。近期还有多个 NPU 相关的修复和优化（如 #27152、#27201），但本 PR 专注于版本升级配套，不涉及逻辑改动。