

PR #25326 完整报告

sgl-project/sglang

chore: bump sgl-kernel version to 0.4.2.post2

合并时间: 2026-05-15 08:53

原文链接: <http://prhub.com.cn/sgl-project/sglang/pull/25326>

执行摘要

- 一句话: 将 sgl-kernel 版本升至 0.4.2.post2
- 推荐动作: 该 PR 为常规维护操作, 无需深入审查。可快速合并。

功能与动机

自动化的版本号更新流程: sglang-bot 提交了版本号变更, Fridge003 手动调整 CI 配置以移除不再需要的 CUDA 12.9 构建, 确保构建流程的简洁和正确性。

实现拆解

1. 更新版本号文件: 修改 sgl-kernel/python/sgl_kernel/version.py, 将 `__version__` 从 "0.4.2.post1" 改为 "0.4.2.post2"。
2. 同步所有平台 pyproject.toml: 分别修改 sgl-kernel/pyproject.toml (CUDA)、pyproject_cpu.toml、pyproject_musa.toml、pyproject_rocm.toml 中的 version 字段, 统一为 "0.4.2.post2"。
3. 移除 CI 中的 CUDA 12.9 构建: 在 .github/workflows/_pr-test-sgl-kernel-build.yml 中删除了 include 列表里 cuda-version: "12.9" 的条目, 共计删除了 2 行。
4. 更新 CMakeLists.txt 中的版本字符串: sgl-kernel/CMakeLists.txt 中的项目版本也从 0.4.2.post1 更新为 0.4.2.post2。

关键文件:

- sgl-kernel/python/sgl_kernel/version.py (模块 内核核心; 类别 source; 类型 core-logic) : 核心版本号定义文件, 确定了 sgl-kernel 的运行时版本, 是本次升级的主要目标。
- sgl-kernel/pyproject.toml (模块 构建配置; 类别 config; 类型 configuration) : CUDA 平台的 pyproject.toml, 定义了项目元数据和构建配置, 版本号需与 version.py 同步。
- .github/workflows/_pr-test-sgl-kernel-build.yml (模块 CI 工作流; 类别 infra; 类型 infrastructure) : CI 构建策略文件, 移除了 CUDA 12.9 的构建项, 确保 CI 不运行已过时的配置。
- sgl-kernel/pyproject_cpu.toml (模块 构建配置; 类别 config; 类型 configuration) : CPU 平台 pyproject.toml, 同步版本号以保持一致。
- sgl-kernel/pyproject_musa.toml (模块 构建配置; 类别 config; 类型 configuration) : MUSA (MThreads) 平台 pyproject.toml, 同步版本号。

- sgl-kernel/pyproject_rocm.toml (模块 构建配置; 类别 config; 类型 configuration) : ROCm (AMD) 平台 pyproject.toml, 同步版本号。
- sgl-kernel/CMakeLists.txt (模块 构建配置; 类别 docs; 类型 documentation) : CMake 构建配置中的项目版本号, 同步更新以确保一致性。

关键符号: 未识别

关键源码片段

[.github/workflows/_pr-test-sgl-kernel-build.yml](#)

CI 构建策略文件, 移除了 CUDA 12.9 的构建项, 确保 CI 不运行已过时的配置。

```
# .github/workflows/_pr-test-sgl-kernel-build.yml (片段)
# 在 matrix.include 中定义了多个构建配置
jobs:
  build:
    strategy:
      matrix:
        include:
          - python-version: "3.10"
            cuda-version: "13.0"
          # 以下条目已被删除:
          # - python-version: "3.10"
          # cuda-version: "12.9"
```

评论区精华

PR 没有 review 评论; Fridge003 在合并前手动添加了两个提交 ([upd](#) 和 [remove cu129 build in ci run](#)), 并在 Issue 评论区使用 [/tag-and-rerun-ci](#) 触发 CI 重跑。这些操作作用意在修复构建环境的一致性, 具体细节未展开讨论。

- 暂无高价值评论线程

风险与影响

- 风险: 本次变更为单纯的版本号更新和 CI 配置清理, 不涉及代码逻辑变更, 因此回归风险极低。唯一可能的风险是: 如果下游依赖通过硬编码版本号或范围约束来引用 sgl-kernel, 则版本号的增加可能导致依赖冲突, 但通常 post2 为 patch 级别, 预期兼容。
- 影响: 对用户无直接影响; 对开发者和 CI 系统, 版本号更新将使得后续新的内核构建能正确发布。移除 CUDA 12.9 构建减少了 CI 矩阵的冗余, 节省了构建资源和时间。
- 风险标记: 低风险例行更新

关联脉络

- 暂无明显关联 PR