

PR #28572 完整报告

sgl-project/sglang

chore: bump sglang-kernel version to 0.4.4

合并时间: 2026-06-18 07:13

原文链接: <http://prhub.com.cn/sgl-project/sglang/pull/28572>

执行摘要

- 一句话: sglang-kernel 版本号从 0.4.3 升至 0.4.4
- 推荐动作: 该 PR 属于常规依赖更新, 无需深入审查。建议合并前确认 CI 测试通过, 确保新内核版本与当前代码兼容。

功能与动机

sgl-kernel/pyproject.toml 中的内核版本已更新为 0.4.4, 需要同步更新 SGLang 项目中所有引用 sglang-kernel 版本的地方, 以保持依赖版本一致。

实现拆解

1. 更新 `python/pyproject.toml`: 将 sglang-kernel 依赖版本从 `==0.4.3` 改为 `==0.4.4`。
2. 更新 `python/sglang/srt/entrypoints/engine.py`: 在版本检查逻辑中, 将 `assert_pkg_version` 的预期版本从 "0.4.3" 改为 "0.4.4"。
3. 更新 `docker/Dockerfile`: 将构建参数 `SGL_KERNEL_VERSION` 从 0.4.3 改为 0.4.4, 确保 Docker 镜像构建时使用正确版本的内核。

关键文件:

- `python/sglang/srt/entrypoints/engine.py` (模块 入口引擎; 类别 source; 类型 core-logic) : 源码中执行版本检查的核心位置, 确保启动时 sglang-kernel 版本正确。
- `python/pyproject.toml` (模块 依赖管理; 类别 config; 类型 configuration) : 项目依赖声明文件, 版本号变更直接影响包安装。
- `docker/Dockerfile` (模块 Docker 构建; 类别 infra; 类型 infrastructure) : Docker 镜像构建时使用该参数安装指定版本内核。

关键符号: 未识别

关键源码片段

`python/sglang/srt/entrypoints/engine.py`

源码中执行版本检查的核心位置, 确保启动时 sglang-kernel 版本正确。

```
# python/sglang/srt/entrypoints/engine.py ( 第 1289-1294 行 )
if _is_cuda:
    assert_pkg_version(
```

```
"sglang-kernel",  
"0.4.4", # 从 0.4.3 升级为 0.4.4  
"Please reinstall the latest version with `pip install sglang-kernel --force-reinstall`",  
)
```

评论区精华

无 review 讨论，仅有一个来自 gemini-code-assist 的自动评论，提示已达到每日配额限制。

- 暂无高价值评论线程

风险与影响

- 风险：版本升级是最小化的补丁版本更新（0.4.3 -> 0.4.4），预期兼容。但仍需关注以下几点：
 - 若新版内核存在未预见的行为变化，可能导致推理结果差异。
 - Dockerfile 中的版本更新确保镜像构建使用新内核，但未测试所有模型。
- 影响：影响范围：本项目所有使用 sglang-kernel 的场景，包括推理、训练等。影响程度：低——仅版本号变更，无 API 或行为改动，升级为兼容性补丁。
- 风险标记：缺少测试覆盖

关联脉络

- PR #28556 chore: bump sgl-kernel version to 0.4.4: 本 PR 是 28556 的后续，28556 更新了 sgl-kernel 自身版本，本 PR 同步更新 SGLang 中对该版本的引用。