

PR #31618 完整报告

sgl-project/sglang

chore: bump sglang-kernel version to 0.4.5

合并时间: 2026-07-18 06:58

原文链接: <http://prhub.com.cn/sgl-project/sglang/pull/31618>

执行摘要

- 一句话: 同步 sglang-kernel 版本至 0.4.5
- 推荐动作: 建议关注 CI 结果, 确认所有测试通过后再合并。本身无需精读, 可作为自动化依赖更新的参考案例。

功能与动机

版本同步: sgl-kernel/pyproject.toml 中定义的 kernel 版本已更新为 0.4.5, 需要在 SGLang 的其他文件中保持一致的版本引用, 确保依赖正确安装和运行时的版本检查通过。

实现拆解

1. 更新依赖声明: 在 python/pyproject.toml 中将 sglang-kernel 的版本约束从 ==0.4.4 改为 ==0.4.5, 确保 pip install 时拉取正确的版本。
2. 更新运行时版本检查: 在 python/sglang/srt/entrypoints/engine.py 的 _set_envs_and_config 函数中, 将 assert_pkg_version 调用的期望版本从 "0.4.4" 改为 "0.4.5", 使得 CUDA 环境下启动时会校验 sglang-kernel 版本是否为 0.4.5。
3. 更新 Docker 构建参数: 在 docker/Dockerfile 中将构建参数 SGL_KERNEL_VERSION 从 0.4.4 改为 0.4.5, 确保 Docker 镜像构建时安装正确的 kernel 版本。

关键文件:

- python/sglang/srt/entrypoints/engine.py (模块 运行时; 类别 source; 类型 core-logic) : 运行时版本检查的核心位置, 变更确保 CUDA 环境下启动时校验 sglang-kernel 版本为 0.4.5
- python/pyproject.toml (模块 依赖配置; 类别 config; 类型 configuration) : 项目依赖声明, pip 安装时的版本约束更新
- docker/Dockerfile (模块 部署脚本; 类别 infra; 类型 infrastructure) : Docker 镜像构建时安装 kernel 版本的构建参数

关键符号: _set_envs_and_config

关键源码片段

python/sglang/srt/entrypoints/engine.py

运行时版本检查的核心位置, 变更确保 CUDA 环境下启动时校验 sglang-kernel 版本为 0.4.5

```
# python/sglang/srt/entrypoints/engine.py 中版本检查部分
if not get_bool_env_var("SGLANG_SKIP_SGL_KERNEL_VERSION_CHECK"):
    if server_args.attention_backend == "flashinfer":
        assert_pkg_version(
            "flashinfer_python",
            "0.6.15",
            "Please uninstall the old version and "
            "reinstall the latest version by following the instructions "
            "at https://docs.flashinfer.ai/installation.html.",
        )
    if _is_cuda:
        assert_pkg_version(
            "sglang-kernel",
            "0.4.5", # 版本号从 0.4.4 更新为 0.4.5
            "Please reinstall the latest version with `pip install sglang-kernel --force-reinstall`,
        )
```

评论区精华

无 review 讨论。只有机器人自动化评论（已被标记 sunset）和合并者 Fridge003 触发的 `/rerun-failed-ci` 命令，表明 CI 有失败项被重新运行。

- 暂无高价值评论线程

风险与影响

- 风险：风险极低。变更仅为版本号更新，不涉及逻辑改动。但若 0.4.5 版本引入了与当前环境不兼容的变更（如 API 变化），可能导致运行时错误。由于该 kernel 是众多功能的基础依赖，升级后需确保 CI 全面通过。
- 影响：影响范围有限但关键：所有使用 sglang-kernel 的部署环境（pip 安装、Docker 镜像）都会安装新版本 0.4.5。对用户透明，但需要验证新 kernel 版本与现有代码的兼容性。
- 风险标记：依赖升级，自动化变更

关联脉络

- PR #31496 chore: bump sgl-kernel version to 0.4.5: 与本 PR 动机相同，均为版本升级，但该 PR 仅更新 sgl-kernel 自身的版本文件，本 PR 同步更新 SGLang 中引用该版本的文件