

PR #27773 完整报告

sgl-project/sglang

[Docs] Fix outdated benchmark marker API in add-jit-kernel skill

合并时间: 2026-06-14 14:47

原文链接: <http://prhub.com.cn/sgl-project/sglang/pull/27773>

执行摘要

- 一句话: 修复 JIT 内核技能中的过时 benchmark API
- 推荐动作: 此 PR 是纯文档更新, 维护技能教程与 API 的一致性。建议合并, 无其他关注点。

功能与动机

SKILL.md 中 Step 5 记录的 benchmark marker API 已过时, 装饰器名称已重命名, 且新增了参数, 教程与真实代码不再匹配。

实现拆解

1. 替换装饰器名称: 将 Step 5 中所有 @marker.mark_benchmark 实例替换为 @marker.benchmark, 将 @marker.mark_args(name, vals) 替换为 @marker.parametrize(names, vals, ci_vals=None)。
2. 修正装饰器顺序说明: 明确指出 @marker.benchmark 是最内层装饰器 (紧贴函数定义), 而非最外层, 以匹配所有树内 benchmark 的写法。
3. 更新 parametrize 参数说明: 说明 names 可接受逗号分隔的相关元组轴 (如 "h,d"), 并新增 ci_vals 参数作为内置 CI 缩简机制。
4. 添加 memory_output 旋钮文档: 说明该参数默认值为 "out", 对于返回 None 的就地内核是必需的。
5. 刷新示例和参考描述: 更新了 bench_scale.py 示例以及 bench_qknorm.py / bench_store_cache.py 的参考描述。
6. 验证其余部分: 确认 Steps 0-4、C++ 抽象参考、Python 包装器 API、CI 套件名称和工作流引用均与当前代码库一致, 未作更改。

关键文件:

- .claude/skills/add-jit-kernel/SKILL.md (模块 技能文档; 类别 docs; 类型 documentation)
: 该文件是唯一变更的文件, 更新了过时的 benchmark API 文档以匹配当前实现。

关键符号: 未识别

评论区精华

无人工 review 评论。仅有一个由 [gemini-code-assist\[bot\]](#) 自动生成的评论, 总结了变更内容, 未提供额外反馈。

- 暂无高价值评论线程

风险与影响

- 风险：无风险。此 PR 仅修改了技能文档中的 Markdown 内容，未涉及任何运行时代码、配置文件或测试。文档准确性已通过对树内 benchmark 和 marker.py 当前代码的验证得到保证。
- 影响：影响范围仅限于使用 add-jit-kernel 技能的开发者，确保他们获得正确的 benchmark 写法指导。无用户或系统影响。
- 风险标记：暂无

关联脉络

- PR #28165 Unify NVTX annotation helpers and split the enable gate per subsystem: 同属 observability / documentation 改进，但此 PR 更关注 API 文档同步。
- PR #28169 [diffusion] UX: reduce attention backend log noise: 同为降低噪声 / 提升开发者体验的文档类变更。