

# PR #20276 完整报告

sgl-project/sglang

[HiCache] fix: graceful shutdown of pending async tasks in bench\_mix.py

合并时间: 2026-03-29 15:46

原文链接: <http://prhub.com.cn/sgl-project/sglang/pull/20276>

## 执行摘要

本次 PR 修复了 `bench_mix.py` benchmark 脚本在成功运行后输出误导性错误消息的问题，通过引入异步任务优雅关闭机制，实现了干净的退出，提升用户体验和代码可维护性。

## 功能与动机

当 `bench_mix.py` benchmark 完成时，由于 event loop 提前关闭而 pending 的异步任务被强制销毁，导致输出大量 "Event loop is closed" 和 "Task was destroyed but it is pending" 错误消息，尽管 benchmark 实际成功。这混淆了用户，因此需要修复以提供清晰的输出，避免误导性错误干扰性能评估。

## 实现拆解

仅修改了 `benchmark/hicache/bench_mix.py` 文件中的 `request_loop` 函数：

- 在函数内部添加 `tasks` 列表，用于跟踪所有通过 `asyncio.create_task` 创建的任务。
- 修改任务创建逻辑：将 `asyncio.create_task(self.handle_request(new_request))` 改为 `task = asyncio.create_task(self.handle_request(new_request)); tasks.append(task)`。
- 在循环退出时，添加代码取消所有任务并等待完成：

## 评论区精华

review 中仅有一次评论，来自 `gemini-code-assist[bot]`：

"The implementation is sound and follows best practices for asyncio task management."

没有其他讨论或争议，变更被迅速批准，表明实现方案得到认可。

## 风险与影响

- 风险：低。异步任务取消可能引入异常处理问题，但使用 `return_exceptions=True` 减少了风险；只影响 benchmark 关闭逻辑，不涉核心路径，回归风险小。
- 影响：正面。用户获得更干净的 benchmark 输出，团队代码库增加了异步任务管理的范例，有助于提高其他类似场景的代码质量。

## 关联脉络

从历史 PR 看，PR #21294 也涉及 benchmark 脚本的修复，两者都旨在提高测试工具的可靠性。这表明团队持续优化 benchmark 基础设施以支持更稳定的性能评估，未来可能在其他 benchmark 文件中推广类似优雅关闭模式。