

PR #27984 完整报告

sgl-project/sglang

[lint] Enable Ruff UP037 to drop redundant quoted annotations

合并时间: 2026-06-12 08:38

原文链接: <http://prhub.com.cn/sgl-project/sglang/pull/27984>

执行摘要

- 一句话: 启用 Ruff UP037 全库清理冗余引号注解
- 推荐动作: 建议阅读 `.pre-commit-config.yaml` 的配置变更, 了解如何加入新 lint 规则。变更本身无需深入 code review。

功能与动机

启用 Ruff 的 UP037 规则, 自动检测并移除冗余的引号注解, 消除代码噪声, 统一代码风格。详见 PR Body。

实现拆解

1. 修改 `.pre-commit-config.yaml`: 在 Ruff 钩子的 `--select` 中添加 UP037, 使后续每次 pre-commit 都会自动检测并修复冗余引号。
2. 手动运行 pre-commit 钩子 (或等价命令) 执行全库自动修复, 覆盖 184 个源文件。
3. 部分签名因去除引号后长度缩短, 被 black 重新格式化为单行。无需测试或其他配套修改。

关键文件:

- `.pre-commit-config.yaml` (模块 代码风格; 类别 config; 类型 configuration): 启用 UP037 规则的核心配置变更, 使该规则在预提交中生效。
- `python/sglang/srt/disaggregation/decode_hicache_mixin.py` (模块 解耦缓存; 类别 source; 类型 lint-fix; 符号 `_build_decode_prefix_match`, `init`, `_clean_hicache_prefetch_resources`, `_try_hicache_queue_load_back`): 自动修复的源文件之一, 演示了类方法注解引号的去除。
- `python/sglang/srt/disaggregation/common/staging_handler.py` (模块 分段传输; 类别 source; 类型 lint-fix; 符号 `create`, `register_decode_req`, `is_done`, `advance_scatter`): 自动修复的源文件之一, 展示了工厂方法和注册函数的注解引号去除。
- `python/sglang/srt/hardware_backend/mlx/scheduler_mixin.py` (模块 MLX 调度; 类别 source; 类型 lint-fix; 符号 `_finalize_mlx_pending_job`, `event_loop_overlap_mlx`, `_launch_fresh`): 自动修复的源文件之一, 展示了 MLX 调度器 mixin 中的注解引号去除。

关键符号: `_build_decode_prefix_match`, `create`, `event_loop_overlap_mlx`, `register_decode_req`, `parse`

关键源码片段

python/sglang/srt/disaggregation/decode_hicache_mixin.py

自动修复的源文件之一，演示了类方法注解引号的去除。

```
def _build_decode_prefix_match(self, req: Req, result: Any) -> DecodePrefixMatch:
    # 此函数原本签名为 req: "Req", 因启用 UP037 后不再需要引号
    # (文件顶部包含 from __future__ import annotations)
    prefix_indices = result.device_indices
    l1_prefix_len = len(prefix_indices)
    l2_host_hit_length = result.host_hit_length

    l3_storage_hit_length = 0
    last_host_node = None
    if self.scheduler.enable_decode_hicache:
        last_host_node = result.last_host_node
        if last_host_node.backuped or last_host_node is self.tree_cache.root_node:
            matched_len = l1_prefix_len + l2_host_hit_length
            suffix_tokens = req.origin_input_ids[matched_len:]
            last_hash = last_host_node.get_last_hash_value()
            prefix_keys = (
                last_host_node.get_prefix_hash_values(last_host_node.parent)
                if self.tree_cache.hicache_storage_pass_prefix_keys
                else None
            )
            l3_storage_hit_length = self.tree_cache.query_storage_hit_length(
                last_host_node, suffix_tokens, last_hash, prefix_keys,
            )

    return DecodePrefixMatch(
        prefix_indices=prefix_indices,
        l2_host_hit_length=l2_host_hit_length,
        l3_storage_hit_length=l3_storage_hit_length,
        last_device_node=result.last_device_node,
        last_host_node=last_host_node if l3_storage_hit_length > 0 else None,
    )
```

评论区精华

无 review 讨论。作者自行检查了字节码编译通过且 pre-commit 三件套 (ruff, black, isort) 达到不动点后合并。

- 暂无高价值评论线程

风险与影响

- 风险：风险极低。UP037 仅在可证明为冗余时去除引号，且对 from __future__ import annotations 环境或函数作用域注解生效。但跨越 185 个文件，可能引入合并冲突，尤其当其他分支正修改同一文件的注解区域。

- 影响：影响所有开发者：从此提交开始，新增代码若使用 `from __future__ import annotations` 则不需要再写注解引号；pre-commit 自动清理遗留引号。CI 时间无影响。
- 风险标记：跨 185 文件改动，合并冲突风险

关联脉络

- 暂无明显关联 PR