

PR #28627 完整报告

sgl-project/sglang

[UnifiedTree]: move some kl test from base to extra stage

合并时间: 2026-06-18 16:39

原文链接: <http://prhub.com.cn/sgl-project/sglang/pull/28627>

执行摘要

- 一句话: 将 4 个 KL 测试从 base 移至 extra 阶段
- 推荐动作: 该 PR 为纯粹的 CI 配置调整, 无需深入阅读。

功能与动机

将执行时间较长或稳定性要求较高的 KL 测试从 base CI 阶段移到 extra CI 阶段, 以优化 base CI 的运行时间和稳定性。

实现拆解

变更仅修改 4 个测试文件中的 `register_cuda_ci` 调用参数:

1) 将 `stage` 从 "base-c" 改为 "extra-b"; 2) 根据历史执行时间中位数更新 `est_time` 参数。

关键文件:

- `test/registered/radix_cache/unified_radix_tree/test_unified_radix_cache_kl_cp.py` (模块 测试配置; 类别 test; 类型 test-coverage) : 从 base-c 移至 extra-b, `est_time` 从 400 改为 950
- `test/registered/radix_cache/unified_radix_tree/test_unified_radix_cache_kl_dsv4.py` (模块 测试配置; 类别 test; 类型 test-coverage) : 从 base-c 移至 extra-b, `est_time` 从 1200 改为 1000
- `test/registered/radix_cache/unified_radix_tree/test_unified_radix_cache_kl_mamba.py` (模块 测试配置; 类别 test; 类型 test-coverage) : 从 base-c 移至 extra-b, `est_time` 从 768 改为 800
- `test/registered/radix_cache/unified_radix_tree/test_unified_radix_cache_hicache_pp_kl.py` (模块 测试配置; 类别 test; 类型 test-coverage) : `est_time` 从 400 改为 900, 仍保留在 base-c

关键符号: 未识别

评论区精华

无 review 争议或讨论, 仅由 `gemini-code-assist[bot]` 自动评论指出变更内容, 无进一步反馈。

- 暂无高价值评论线程

风险与影响

- 风险：无技术风险。仅涉及 CI 阶段配置变更，不影响源码功能。
- 影响：影响仅限于 CI 流水线的测试调度：这四个 KL 测试将不再在 base CI 中运行，而是仅在 extra CI 中执行，可能减低 base CI 耗时。
- 风险标记：暂无

关联脉络

- PR #27291 [HiCache] Fix SWA L3 cache miss due to a prefetch/hit len mismatch: 修改了部分相同的测试文件，涉及 KL 测试