PR #22563 完整报告

sgl-project/sglang

fix: match est_time updates by backend, not just suite

合并时间：2026-04-11 08:54

原文链接：http://prhub.com.cn/sgl-project/sglang/pull/22563

执行摘要

- 一句话：修复 CI 测试时间估算脚本，按后端硬件区分时间统计，避免跨后端数据污染。
- 推荐动作：该 PR 虽小但展示了 CI 基础设施中一个重要的数据隔离问题。建议精读以理解：1) 如何通过数据结构设计避免数据污染；2) 正则表达式在配置更新中的精确匹配技巧。对于负责 CI 维护的工程师，这是值得参考的修复模式。

功能与动机

PR body 明确指出：在 PR #22561 中，test_eagle3_basic.py 文件同时包含 register_cuda_ci 和 register_amd_ci 注册，且使用相同的测试套件 (stage-b-test-1-gpu-small)。原脚本使用 register_\w+_ci 正则表达式匹配任何后端，导致 CUDA 的中位时间被错误应用到 AMD 注册上，将 est_time 从 50 改为 70。这暴露了跨后端时间数据污染的问题，需要修复以确保每个后端只使用自己的性能数据。

实现拆解

修改集中在 scripts/ci/update_est_time.py 文件：

1. 数据结构变更：将时间收集字典的键从 (relative_path, suite) 改为 (relative_path, suite, backend)，后端通过 determine_backend 函数从作业名中提取。
2. 中位数计算：compute_medians 函数相应更新，处理三元组键。
3. 分组逻辑：update_est_times 函数中将 by_file 从 {rel_path: {suite: median}} 改为 {rel_path: [(suite, backend, median), ...]}。
4. 正则表达式：将匹配模式从 register_\w+ci 改为 register{backend}_ci，确保只匹配对应后端的注册调用。

关键文件：

- scripts/ci/update_est_time.py（模块 CI 基础设施）：唯一修改的文件，包含时间收集、计算和更新的核心逻辑变更。

关键符号：collect_timings, compute_medians, update_est_times

评论区精华

无 review 评论，PR 由作者直接合并。从提交信息看，作者清晰描述了问题根源和修复方案。

- 暂无高价值评论线程

风险与影响

- 风险：技术风险较低：
  1. 回归风险：修改仅影响 CI 测试时间估算逻辑，不涉及核心推理路径，但若正则表达式匹配错误可能导致 est_time 更新到错误行。
  2. 兼容性：需要确保 determine_backend 函数能正确从所有 CI 作业名中提取后端标识，否则可能遗漏某些后端的时间数据。
  3. 数据完整性：修复后每个后端将独立统计时间，但若某个后端数据点不足 (MIN_DATA_POINTS)，其 est_time 可能不会被更新。
- 影响：影响范围有限但重要：
  1. 对用户：无直接影响，仅影响内部 CI 基础设施。
  2. 对系统：确保 CI 测试时间估算更准确，有助于优化测试负载均衡和资源调度。
  3. 对团队：修复了跨后端时间数据污染问题，避免未来类似 PR #22561 中的错误更新，提升 CI 配置的可靠性。
- 风险标记：正则表达式匹配风险，后端标识提取依赖

关联脉络

- PR #22557 fix: track est_time per suite instead of per backend: 同样修改 update_est_time.py，但方向相反：该 PR 将时间统计从按后端改为按测试套件，而本 PR 是在此基础上进一步细化，需要区分同一套件下的不同后端。两者共同构成 CI 时间统计的演进脉络。
- PR #22545 feat: add weekly workflow to update CI test est_time values: 同样涉及 update_est_time.py 脚本，添加了自动化工作流。本 PR 修复了该工作流依赖的时间统计逻辑。
- PR #22561 未在历史 PR 列表中，但 PR body 提及：PR body 明确指出 #22561 中出现了因本 bug 导致的错误更新（test_eagle3_basic.py 的 AMD est_time 被 CUDA 数据污染），是本修复的直接诱因。