

PR #28738 完整报告

sgl-project/sglang

[CI] Publish 4-GPU nightly profiler traces

合并时间: 2026-06-20 06:27

原文链接: <http://prhub.com.cn/sgl-project/sglang/pull/28738>

执行摘要

- 一句话: 4-GPU 夜间 CI 新增 trace 发布步骤
- 推荐动作: 变更简单直接, 无需深度 review, 但值得关注的是团队如何通过 CI 自动化实现可见性提升。如果正在配置类似流程, 可参考 `publish_traces.py` 和 `PERFETTO_RELAY_URL` 的使用方式。

功能与动机

在之前其他 NVIDIA 夜间 job 已经支持了 trace 发布和 profiler 链接显示, 但 4-GPU B200 和 GB300 的 job 没有包含该步骤。PR Body 明确指出目的: "Add trace publishing for the 4-GPU B200 and GB300 nightly perf jobs so their profiler links show up like the other NVIDIA nightlies"。

实现拆解

1. 为 `nightly-4-gpu-b200` job 添加 trace 发布: 在 `nightly-test-nvidia.yml` 中, 在 `nightly-4-gpu-b200 suite` 的 "Run test" 步骤后新增 Publish traces to storage repo 步骤。该步骤使用 `always()` 条件确保即使测试失败也执行, 并设置 `continue-on-error: true`。同时, 在 "Run test" 步骤中添加环境变量 `TRACE_BASE_URL`、`PERFETTO_RELAY_URL` 和 `GPU_CONFIG: "4-gpu-b200"`, 供测试和 trace 流程使用。
2. 为 `nightly-4-gpu-gb300` job 添加 trace 发布: 对 `nightly-4-gpu-gb300` job 做完全相同的一组变更: 添加环境变量和发布步骤, 其中 `GPU_CONFIG` 设为 "4-gpu-gb300"。
3. 实现发布逻辑: 发布步骤本身是一个 shell 脚本, 遍历 `test/performance_profiles_*/` 目录下的 trace 目录 (由测试流程产生), 若存在则调用 `python3 scripts/ci/utis/publish_traces.py` 上传, 上传成功后删除所有 `*.json.gz` 本地 trace 文件以节省磁盘空间。若没有 trace 目录则跳过。

关键文件:

- `.github/workflows/nightly-test-nvidia.yml` (模块 CI 配置; 类别 `infra`; 类型 `infrastructure`): 唯一变更文件, 在 `nightly-4-gpu-b200` 和 `nightly-4-gpu-gb300` 两个 job 中新增 trace 发布步骤与环境变量。

关键符号: 未识别

评论区精华

该 PR 没有 review 评论，仅由 Fridge003 直接批准。Gemini Code Assist 因文件类型不支持而未能生成 review。因此无实质性讨论。

- 暂无高价值评论线程

风险与影响

- 风险：

1. 新增 secret 使用：GH_PAT_FOR_NIGHTLY_CI_DATA 用于向存储仓库写入，若该 secret 配置不当或权限不足可能导致发布失败，但已设置 `continue-on-error: true`，不会影响测试本身。
2. 磁盘清理可能误删：`find test/performance_profiles_*/ -name '*.json.gz' -delete` 在发布成功后删除 trace 文件。若发布步骤之前有其他步骤也在同一目录下产生了 `.json.gz` 文件（可能性极低），可能被误删。不过该目录命名约定明确由测试框架生成，风险可控。
 - 影响：影响范围：仅影响 NVIDIA 夜间 CI 的 4-GPU B200 和 GB300 job。影响程度：低。仅新增一个 CI 步骤，不修改任何业务逻辑或测试代码。正确执行后，开发者可以在 CI 报告中看到 profiler 链接，提升调试效率。

- 风险标记：新增 secret 依赖，磁盘清理操作可能误删

关联脉络

- PR #28536 ci: run GB300 nightly suite in the standard Nvidia nightly workflow: 相关 PR，将 GB300 测试加入同一工作流，本 PR 在此基础上添加 trace 发布。