

PR #25387 完整报告

sgl-project/sglang

ci: standalone PR awareness workflow (decouple from pr-test reusable chain)

合并时间: 2026-05-15 18:20

原文链接: <http://prhub.com.cn/sgl-project/sglang/pull/25387>

执行摘要

将 PR 意识 (CI 状态块) 从 `pr-test.yml` 的可复用链中解耦为一个独立的 `pr-states.yml` 工作流, 以消除 `pulls: write` 权限的跨工作流传播, 并修复 `release-branch-cut.yml` 的启动失败。变更涉及删除旧文件、新建独立工作流以及调整两个主测试工作流的权限。

功能与动机

原本 PR 意识块更新嵌在 `pr-test.yml` 和 `pr-test-extra.yml` 的 `call-pr-awareness` job 中, 导致 `pulls: write` 权限沿工作流链传递, 增加了安全风险。同时该设计在 `release-branch-cut.yml` 中引发 `startup_failure`。本 PR 将更新逻辑独立为 `pr-states.yml`, 只监听 `pull_request` 事件, 不再依赖测试链的 `workflow_call`, 从而实现了权限隔离和稳定性提升。

实现拆解

1. 删除旧有工作流: 移除 `.github/workflows/_pr-awareness-comment.yml` (123 行), 该文件通过 `workflow_call` 机制被调用, 具有 `pulls: write` 权限。
2. 新建独立工作流: 创建 `.github/workflows/pr-states.yml` (105 行), 基于 `pull_request` 的 `opened`、`synchronize`、`reopened`、`labeled`、`unlabeled` 事件触发; 通过 GitHub API 获取 `pr-test.yml` 和 `pr-test-extra.yml` 的最新运行, 并包含 `findRunWithRetry` 重试逻辑 (最多 6 次, 间隔 5s) 以及 `isReal` 跳过完全 `skipped` 的运行; 使用 `concurrency` 组串行化同一 PR 的多次更新。
3. 调整主测试工作流: 在 `pr-test.yml` 和 `pr-test-extra.yml` 中删除 `call-pr-awareness` job, 并将 `permissions.pull-requests` 从 `write` 降级为 `read`, 彻底切断权限提升路径。
4. Commit 演进的精炼: 5 个 commit 依次处理移动、跳过优化、分隔符去重、重命名、清理一次性迁移代码, 最终版本干净明了。

`.github/workflows/pr-states.yml`

新增的独立工作流, 替代旧 `awareness` 功能, 通过 `pull_request` 事件触发, 包含重试和 `skipped` 过滤。

```
/* pr-states.yml - 核心 update-pr-body 步骤的 Script */
const sha = context.payload.pull_request.head.sha;
const labels = context.payload.pull_request.labels.map(l => l.name);
const hasExtra = labels.includes('run-ci-extra');
```

```
// 按 workflow 文件名查找最近一次运行 (最多重试 6 次, 间隔 5 秒)
```

```

async function findRunBySha(workflowFile) {
  const { data } = await github.rest.actions.listWorkflowRuns({
    owner: context.repo.owner,
    repo: context.repo.repo,
    workflow_id: workflowFile,
    head_sha: sha,
    per_page: 1,
  });
  return data.workflow_runs[0] || null;
}

async function findRunWithRetry(workflowFile, maxAttempts = 6, delayMs = 5000) {
  for (let i = 0; i < maxAttempts; i++) {
    const run = await findRunBySha(workflowFile);
    if (run) return run;
    if (i < maxAttempts - 1) await new Promise(r => setTimeout(r, delayMs));
  }
  return null;
}

// 跳过被完全跳过 (skipped) 的运行 (如未触发 label 时)
const isReal = (run) => run && run.conclusion !== 'skipped';

// 查找 pr-test 和 pr-test-extra 的运行结果
const ptRun = await findRunWithRetry('pr-test.yml');
const peRun = hasExtra ? await findRunWithRetry('pr-test-extra.yml') : null;

const ptText = isReal(ptRun) ? `[Run #${ptRun.id}](${ptRun.html_url})` : '_Not run yet_';
const peText = !hasExtra
  ? ':warning: **Not enabled** — add `run-ci-extra` label to opt in.'
  : (isReal(peRun) ? `[Run #${peRun.id}](${peRun.html_url})` : ':warning: **Not run on latest push**');

// 构建新状态块并替换 PR body
const newBlock = [
  '<!-- pr-states:start -->',
  '___',
  '### CI States',
  '',
  `Latest PR Test: ${ptStart}${ptText}${ptEnd}`,
  `Latest PR Test (Extra): ${peStart}${peText}${peEnd}`,
  '<!-- pr-states:end -->',
].join('\n');

```

评论区精华

本 PR 无 reviewer 讨论，仅有一条 bot 自动回复和一条 `/tag-and-rerun-ci` 命令。作者通过多次 commit 自审查直至合入，未出现设计争议。

风险与影响

- 权限降级: pr-test.yml 和 pr-test-extra.yml 的 pull-requests 权限从 write 降为 read, 若其他未发现的 job 依赖写权限可能失效, 但当前检查未发现此类情况。
- API 依赖: pr-states.yml 依赖 GitHub API 查询运行, 若限速或不可用, 状态块会延迟显示, 但重试机制缓解了短暂故障。
- 并发写入: 独立工作流使用 concurrency 组确保同一 PR 的更新串行化, 避免状态相互覆盖。
- 影响范围: 仅限 CI 配置, 不影响业务代码; 开发者无直接感知, PR 状态块继续正常显示。

关联脉络

本 PR 是对 #25320 和 #25322 的跟进, 前序 PR 引入了定时调度和废弃 `/rerun-stage`, 本 PR 在此基础上进一步完善了权限模型和解耦。整体方向是将 CI 权限收缩、职责明确化, 为后续的安全加固和运维简化打下基础。