

PR #25320 完整报告

sgl-project/sglang

ci: dispatch pr-test-extra.yml from pr-test.yml on the scheduled cron

合并时间: 2026-05-15 16:01

原文链接: <http://prhub.com.cn/sgl-project/sglang/pull/25320>

执行摘要

- 一句话: 定时调度触发额外测试层, 引入 PR awareness 注释
- 推荐动作: 值得精读该 PR 的工作流设计, 尤其是 `_pr-awareness-comment.yml` 中通过 HTML 注释槽位替换和 `concurrency group` 序列化的实践, 以及 `call-pr-test-extra` 的调度策略。对于维护多层次 CI 的仓库有参考价值。

功能与动机

现有每天 3 次的定时调度仅覆盖基础测试层 (`pr-test.yml`), 而 `extra-tier` 测试 (包括 #24725 中移到 `run-ci-extra` 标签门控下的约 50 个测试, 以及 #25203 和 #25236 中添加的 B200/H200 8 GPU 条件测试) 仅通过 PR 标签或手动 `workflow_dispatch` 触发, 缺少定时回归兜底。

实现拆解

1. 新增 `_pr-awareness-comment.yml` 可重用工作流: 通过 `workflow_call` 接收 `workflow_kind` 参数, 在 PR body 中插入 / 更新两个槽位 (`pr-test` 和 `pr-test-extra`), 显示对应 CI 的运行状态。使用 `actions/github-script` 和 `pulls.get/update` API, 通过 `concurrency group` 序列化并发写操作。
2. 修改 `pr-test.yml`: 添加 `run_all_tests` 输入参数; 新增 `call-pr-test-extra` 作业, 仅在 `schedule` 事件或 `test_parallel_dispatch=true` 时触发, 调用 `pr-test-extra.yml`; 新增 `call-pr-awareness` 作业, 在 `check-changes` 后立即更新意识块。
3. 修改 `pr-test-extra.yml`: 添加 `call-pr-awareness` 作业, 确保即使 PR 没有 `run-ci-extra` 标签 (作业被跳过) 也能更新意识块, 让作者看到此次运行是空操作。
4. 调整健康检查跳过条件: 在 `SKIP_STAGE_HEALTH_CHECK` 中增加 `test_parallel_dispatch` 和 `run_all_tests` 条件, 避免批量 / 全量运行时因单个阶段失败而提前终止。
5. 为 Stage Job 添加名称: 在 `_pr-test-stage.yml` 和 `_pr-test-check-changes.yml` 中添加 `name` 字段, 使 GHA UI 中的矩阵任务显示为“extra-a-test-1-gpu-small (0)”而非裸“(0)”, 提升可读性。

关键文件:

- `.github/workflows/_pr-awareness-comment.yml` (模块 CI 工作流; 类别 `infra`; 类型 `infrastructure`; 符号 `existingBlock`, `extractSlot`, `newPrTest`, `newPrExtra`): 新增的核心

workflows, 负责在 PR body 中维护 CI 意识块, 实现两个工作流的运行状态展示。

- .github/workflows/pr-test.yml (模块 CI 工作流; 类别 infra; 类型 infrastructure) : 主 CI 工作流, 新增 run_all_tests 输入和 call-pr-test-extra 作业, 实现调度时触发 extra 测试。
- .github/workflows/pr-test-extra.yml (模块 CI 工作流; 类别 infra; 类型 infrastructure) : extra 测试工作流, 新增 call-pr-awareness 作业, 确保即使 PR 无标签也能更新意识块。
- .github/workflows/_pr-test-check-changes.yml (模块 CI 工作流; 类别 infra; 类型 infrastructure) : 可重用工作流, 增加 name 字段使 GHA UI 显示更清晰。
- .github/workflows/_pr-test-stage.yml (模块 CI 工作流; 类别 infra; 类型 infrastructure) : 可重用工作流, 增加 name 字段使矩阵任务名称包含 stage 名, 提升可读性。

关键符号: existingBlock, extractSlot, newPrTest, newPrExtra

关键源码片段

.github/workflows/_pr-awareness-comment.yml

新增的核心工作流, 负责在 PR body 中维护 CI 意识块, 实现两个工作流的运行状态展示。

```
# .github/workflows/_pr-awareness-comment.yml
# 可重用工作流: 在 PR body 中维护 CI 意识块
name: PR Awareness Comment

on:
  workflow_call:
    inputs:
      workflow_kind:
        description: "Which workflow is calling: 'pr-test' or 'pr-test-extra'"
        required: true
        type: string

permissions:
  pull-requests: write

jobs:
  update-pr-body:
    if: github.event_name == 'pull_request'
    runs-on: ubuntu-latest
    # 用 concurrency group 防止 pr-test 和 pr-test-extra 同时更新时发生冲突
    concurrency:
      group: pr-awareness-${{ github.event.pull_request.number }}
      cancel-in-progress: false
    steps:
      - name: Update awareness block in PR body
        uses: actions/github-script@v7
        env:
          KIND: ${{ inputs.workflow_kind }}
        with:
```

```
script: |
  const kind = process.env.KIND;
  const { data: pr } = await github.rest.pulls.get({ ... });
  const body = pr.body || '';
  // 提取现有意识块，保留另一个工作流的槽位
  const existingBlock = (body.match(/<!-- pr-awareness:start -->[\s\S]*?<!-- pr-awareness:end -->/) || [''])[0];
  const prevPrTest = extractSlot(existingBlock, '<!-- slot:pr-test:start -->', '<!-- slot:pr-test:end -->');
  const prevPrExtra = extractSlot(existingBlock, '<!-- slot:pr-test-extra:start -->', '<!-- slot:pr-test-extra:end -->');
  // 仅更新调用者自己的槽位，另一个保持原样
  // ... 具体替换逻辑略
```

评论区精华

无 Review 评论；Issue 中作者提供了两个 Actions 运行链接用于验证，合并者使用 [/tag-and-rerun-ci](#) 重新触发 CI。未发现实质性设计争论。

- 暂无高价值评论线程

风险与影响

- 风险：

1. PR body 并发写入：pr-test 和 pr-test-extra 都可能更新同一 PR 的意识块，虽通过 concurrency group 序列化，但若两个工作流同时触发（可能性低），后执行者可能覆盖前者的槽位。已通过逐槽位替换而非整体覆盖缓解。
2. Extra 状态未接入 required check：call-pr-test-extra 的结果未加入 pr-test-finish 的 needs:，因此定时调度下 extra 测试失败不会影响父运行状态（PR 中仍可见），但不会阻止合入。若需严格门禁需后续跟进。
3. 条件覆盖不全：call-pr-test-extra 的触发条件为 github.event_name == 'schedule' || inputs.test_parallel_dispatch == true，但未包含 workflow_dispatch；通过 test_parallel_dispatch 可模拟全量运行，但文档中未明确说明。- 影响：对用户：无直接功能影响。对系统：定时调度会增加 extra 测试层的执行频率（每天 3 次→实际新增约 50 个测试及 B200/H200 任务），但 CI 资源消耗有限。对团队：extra 测试获得定时回归覆盖，减少 main 上引入回归的风险；PR 创建者能通过意识块快速了解 CI 状态，降低沟通成本。影响程度中等，属于 CI 稳定性提升。- 风险标记：PR body 并发写入风险，Extra 状态未接入 required check，定时调度跳过 stage health check

关联脉络

- PR #24725 [ci] Move ~50 tests under run-ci-extra label gate: 该 PR 引入了 extra-test 的标签门控机制，是本次需要为 extra 测试提供定时调度覆盖的直接原因。
- PR #25203 [ci] Add B200 8-gpu conditional test: 引入 B200 8 GPU 条件测试，属于 extra-tier 测试，缺少定时回归。

- PR #25236 [ci] Add H200 8-gpu conditional test: 引入 H200 8 GPU 条件测试, 同样属于 extra-tier 测试。