

PR #24374 完整报告

sgl-project/sglang

[CI] Exclude flaky h20 stage from check-stage-health root cause set

合并时间: 2026-05-05 02:41

原文链接: <http://prhub.com.cn/sgl-project/sglang/pull/24374>

执行摘要

- 一句话: 排除不稳定的 h20 阶段引发级联失败
- 推荐动作: 此 PR 是小幅 CI 可靠性改进, 值得合并。建议关注 h20 运行器 GPU 残留的根本原因, 并考虑更长期的修复。

功能与动机

h20 运行器经常在启动时残留前序工作的 GPU 内存, 且清理脚本无法回收, 导致作业在运行任何测试之前即失败 (如 run 25316228022 所示, 所有 8 个 GPU 均报错“ERROR: memory >=10%”, 且存在 8 个不可杀死的 PID)。此失败并非 PR 引入, 但会被 `check-stage-health` 视为根因, 导致其他阶段 (如 h200、b200、deepep) 快速失败级联。PR 目的是排除此不稳定的作业, 避免影响其他正常阶段。

实现拆解

在 `.github/actions/check-stage-health/action.yml` 中的根因失败过滤逻辑里, 添加一个精确名称匹配 `stage-c-test-8-gpu-h20` 的早期返回 `false`, 从而将其排除在根因集合之外。

1. 新增条件判断: 在遍历作业寻找根因失败时, 首先检查作业名称是否精确等于 `stage-c-test-8-gpu-h20`, 如果是则直接返回 `false` (即不视为根因)。
2. 保留原有逻辑: 修改仅影响 h20 作业, 其余作业 (如 h200、b200、deepep) 仍按原有逻辑处理。
3. 无其他变更: 仅修改此文件, 无需测试、配置或部署配套。

关键文件:

- `.github/actions/check-stage-health/action.yml` (模块 CI; 类别 infra; 类型 infrastructure): 核心变更文件, 添加了排除 h20 作业的精确名称匹配逻辑。

关键符号: 未识别

关键源码片段

`.github/actions/check-stage-health/action.yml`

核心变更文件, 添加了排除 h20 作业的精确名称匹配逻辑。

```
# .github/actions/check-stage-health/action.yml (片段)
# 在根因失败过滤循环中添加如下判断:
```

```
# h20 runners are flaky (dirty GPU state from prior runs); their failures
# should not cascade fast-fail to other stages. Exact match avoids
# accidentally matching the h200 job names.
if (j.name === 'stage-c-test-8-gpu-h20') {
  return false;
}
```

评论区精华

无实质性 review 讨论。仅有的评论来自 [gemini-code-assist\[bot\]](#)，其表示无反馈。

- 暂无高价值评论线程

风险与影响

- 风险：风险极低。改动仅一行（精确名称匹配），且仅影响 CI 控制流。若 h20 作业名称未来更改，需同步更新此文件，否则排除失效。此外，h20 作业失败仍会报告，但不再级联，这可能掩盖某些真正的根因问题（如 h20 特定代码缺陷），但鉴于失败发生在依赖安装之前，风险较小。
- 影响：
 - 用户 / 开发者：h20 阶段失败不再导致其他阶段被跳过，减少 CI 误报和重试。h20 作业本身仍会失败并显示状态。
 - 系统 / CI：减少不必要的级联失败，提高 CI 效率。
 - 团队：降低因不稳定运行器导致的 CI 阻塞，但需关注 h20 运行器清理问题的根本解决方案。
 - 风险标记：依赖精确作业名称，低风险

关联脉络

- 暂无明显关联 PR