

PR #32174 完整报告

sgl-project/sglang

[Tiny]Correct runner for testing deepgemm

合并时间: 2026-07-23 16:13

原文链接: <http://prhub.com.cn/sgl-project/sglang/pull/32174>

执行摘要

修正 DeepGEMM 构建任务中使用的 CI 运行器标签, 将 `1-gpu-small` 改为 `1-gpu-5090`, 以解决运行器可用性問題。这是一个微小的基础设施配置变更。

功能与动机

DeepGEMM 构建流水线需要特定的 GPU 运行器, 之前的 `1-gpu-small` 可能不适用或不可用, 导致 CI 失败。PR 通过更正运行器标签来适配正确的硬件节点。

实现拆解

1. 修改 CI 配置文件: 在 `.github/workflows/release-whl-deepgemm.yml` 中, 将对应 `arch_label: sm120` 的 job 中的 `runner: 1-gpu-small` 改为 `runner: 1-gpu-5090`。

无。变更仅涉及 YAML 配置中的一行值替换。

评论区精华

无 review 讨论。

风险与影响

- 风险: 极低。仅影响 CI 运行器选择, 不涉及代码逻辑。若 `1-gpu-5090` 运行器不可用, 任务可能会排队或失败。
- 影响: 仅限 DeepGEMM 构建流水线, 不影响其他功能或用户。

关联脉络

该 PR 独立, 与近期其他 DeepGEMM 或 CI 配置变更无直接关联。