

PR #48186 完整报告

vllm-project/vllm

[CI] Right-size test-area timeouts from nightly durations

合并时间: 2026-07-10 15:53

原文链接: <http://prhub.com.cn/vllm-project/vllm/pull/48186>

执行摘要

- 一句话: 根据夜间 CI 数据调整所有测试域的超时时间
- 推荐动作: 建议合并。该 PR 采用数据驱动方法精确调整超时, 值得借鉴。未来可增加更频繁的自动调整机制。

功能与动机

PR body 指出: many premerge timeouts are either well below the real runtime (risking spurious timeouts) or heavily over-provisioned (a hung job sits idle a long time before failing), 因此需要根据夜间实际运行数据精确调整超时。

实现拆解

1. 从 nightly full CI build (77252) 获取每个 step 的实际运行时长 (observed = finished_at - started_at), 仅统计成功运行的 step。
2. 对于 sharded 步骤 (parallelism/%N), 使用所有 shard 的 max runtime。
3. 对主步骤应用公式 $\text{round_up_to_5}(\text{observed} + 15)$; 对 AMD 镜像步骤应用 $\text{round_up_to_5}(\text{observed} + 35)$ (额外 +20 缓冲)。
4. 修改 30 个 .yaml 文件中对应 step 的 timeout_in_minutes 字段, 保持其他配置不变。
5. AMD 额外缓冲是由于其运行在 off-AWS 环境, 网络和共享资源波动大, 且 HuggingFace 数据下载缓存问题 (#36951) 仍存在。

关键文件:

- .buildkite/test_areas/engine.yaml (模块 CI 配置; 类别 config; 类型 configuration): 包含核心测试步骤的超时调整 (Engine, e2e Scheduling, e2e Core 等), 体现了主公式和 AMD 公式的应用
- .buildkite/test_areas/kernels.yaml (模块 CI 配置; 类别 config; 类型 configuration): 涉及多个 GPU 内核测试的超时调整, 如 Kernels Core Operation Test 从 75 分钟增加到 120 分钟
- .buildkite/test_areas/lm_eval.yaml (模块 CI 配置; 类别 config; 类型 configuration): 展示了大幅缩减超时的例子 (如 LM Eval Large Models (8xH200) 从 60 分钟降到 50 分钟, AMD 镜像从 180 分钟降到 60 分钟)

- `.buildkite/test_areas/misc.yaml` (模块 CI 配置; 类别 config; 类型 configuration) : 包含 Miscellaneous 组多个步骤的超时调整, 如 V1 Core + KV + Metrics 从 30 分钟增加到 60 分钟
- `.buildkite/test_areas/models_language.yaml` (模块 CI 配置; 类别 config; 类型 configuration) : 语言模型测试组的超时调整, 如 Extended Generation 从 110 分钟降至 65 分钟

关键符号: 未识别

评论区精华

AndreasKaratzas 建议为 AMD 测试组增加额外的 20 分钟缓冲, 理由是 AMD 环境不在 AWS 上, 网络和共享资源波动大, 且 HuggingFace 数据下载缓存问题未解决。khluu 响应已更新, 最终 AMD 步骤的缓冲从 15 分钟增加到 35 分钟。

- 为 AMD 测试组增加额外 20 分钟缓冲 (other): khluu 接受建议, 为所有 AMD 步骤额外增加了 20 分钟缓冲 (总缓冲 35 分钟)。

风险与影响

- 风险: 主要风险是超时调整基于单次 nightly 运行数据, 可能不具完全代表性; 少数 step 的超时被大幅缩短 (例如从 120 分钟降到 45 分钟), 极端情况下可能仍不足。但公式提供了缓冲, 且 AMD 有额外缓冲, 风险较低。需持续监控 CI 稳定性。
- 影响: 对用户无直接影响。对 CI 系统: 减少因超时设置过短导致的误报失败, 提高 CI 可靠性; 同时减少挂起作业的等待时间 (对于之前超时设置过长的 step)。对团队: CI 更稳定, 减少不必要的重试, 提升开发效率。
- 风险标记: 依赖单次 nightly 观测数据, AMD 环境波动性, 部分步骤超时大幅缩减

关联脉络

- PR #36951 HuggingFace data download caching issue: AMD 额外缓冲的原因之一, 该 issue 仍在开放
- PR #48161 [CI] Increase extract hidden states TP2 timeout: 类似超时调整, 但针对单个测试