

PR #48418 完整报告

vllm-project/vllm

[XPU][CI]Adjust timeout_in_minutes in Intel GPU CI

合并时间: 2026-07-13 23:11

原文链接: <http://prhub.com.cn/vllm-project/vllm/pull/48418>

执行摘要

- 一句话: 调整 Intel GPU CI 超时时间
- 推荐动作: 该 PR 为常规 CI 维护, 无需精读。关注点在于后续是否需进一步优化测试性能以降低超时需求。

功能与动机

根据 PR Body 描述, 目的是 "Adjust timeout_in_minutes in Intel GPU CI", 推测由于 Intel GPU 测试环境性能或负载原因, 原有 30 分钟超时不足以完成测试, 导致 CI 频繁失败。

实现拆解

1. 调整 `test-intel.yaml` (两个位置): XPU example Test 超时从 30 分钟改为 50 分钟, XPU V1 test 从 30 分钟改为 70 分钟, XPU server test 从 30 分钟改为 45 分钟。
2. 调整 `models_multimodal_intel.yaml`: 多个多模态模型测试步骤的超时从 45 分钟调整为 60-70 分钟, 其中 Multi-Modal Processor 从 45 分钟改为 60 分钟。
3. 调整 `misc_intel.yaml`: 多个杂项测试步骤超时从 30 分钟改为 45-90 分钟, 其中 V1 Sample + Logits 从 30 分钟改为 90 分钟 (增幅最大)。
4. 调整 `models_distributed_intel.yaml`: 分布式模型测试从 50 分钟改为 65 分钟。
5. 调整 `basic_correctness.yaml`: XPU Sleep Mode 测试从 30 分钟改为 45 分钟。
6. 调整 `lora_intel.yaml`: 对应步骤超时从 30 分钟改为 45 分钟 (该文件未提供完整上下文, 但模式一致)。

关键文件:

- `.buildkite/hardware_tests/intel_xpu_ci/test-intel.yaml` (模块 CI 配置; 类别 test; 类型 configuration): 硬件测试 CI 配置, 调整了 3 个测试步骤的超时时间 (30→50/70/45 分钟), 是本次变更中唯一的 test 类文件。
- `.buildkite/intel_jobs/models_multimodal_intel.yaml` (模块 CI 配置; 类别 config; 类型 configuration): 多模态模型测试配置, 调整了 4 个步骤的超时时间 (45→70/65/60 分钟), 影响面较广。
- `.buildkite/intel_jobs/misc_intel.yaml` (模块 CI 配置; 类别 config; 类型 configuration): 杂项测试配置, 调整了 5 个步骤的超时时间 (30→45/90/50/45/55 分钟), 其中 V1 Sample + Logits 增幅最大 (30→90), 可能反映该测试负载较重。

- `.buildkite/intel_jobs/test-intel.yaml` (模块 CI 配置; 类别 config; 类型 configuration) : 常规测试配置, 调整了 3 个步骤的超时时间 (30→50/70/45 分钟), 与 `hardware_tests` 下的同名文件变更一致。
- `.buildkite/intel_jobs/models_distributed_intel.yaml` (模块 CI 配置; 类别 config; 类型 configuration) : 分布式测试配置, 调整了分布式模型测试的超时时间 (50→65 分钟)。
- `.buildkite/intel_jobs/basic_correctness.yaml` (模块 CI 配置; 类别 config; 类型 configuration) : 正确性测试配置, 调整了 XPU Sleep Mode 测试的超时时间 (30→45 分钟)。
- `.buildkite/intel_jobs/lora_intel.yaml` (模块 CI 配置; 类别 config; 类型 configuration) : LoRA 相关测试配置, 超时时间从 30 改为 45 分钟。

关键符号: 未识别

评论区精华

无技术讨论。仅有的 review 来自 `claude[bot]` 自动评论 (由于是 fork PR, 自动审核被禁用), 以及 `jikunshang` 的批准。

- 暂无高价值评论线程

风险与影响

- 风险: 无技术风险。仅调整 CI 超时时间, 不影响运行时功能。但过长的超时可能掩盖测试死锁或性能退化问题, 建议配合测试失败日志监控。
- 影响: 直接影响 Intel GPU CI 的稳定性, 减少因超时导致的误报失败。间接影响是 CI 总耗时可能增加, 但并行度未变, 对开发体验影响有限。
- 风险标记: 无风险

关联脉络

- PR #48394 [CI][2/N] reduce CI time: 同一仓库的 CI 时间优化, 与本 PR 相反方向 (减少 vs 增加超时), 反映 CI 稳定性与效率的平衡。