

PR #46418 完整报告

vllm-project/vllm

[ROCm][CI] Purging away redundant test group definitions

合并时间: 2026-06-23 06:42

原文链接: <http://prhub.com.cn/vllm-project/vllm/pull/46418>

执行摘要

- 一句话: 清理 ROCm CI 冗余测试组并优化硬件分配
- 推荐动作: 该 PR 为常规基础设施清理, 变更直接、风险低, 值得快速合并。对于关注 ROCm CI 效率的团队, 可以了解硬件分配策略。

功能与动机

MI250 集群上运行着一些本应由更现代 AMD 硬件 (如 MI300) 执行的测试组, 同时为了缓解 MI250 和 MI355 集群的压力, 需要将部分测试组迁移到 MI300 集群, 并清理冗余定义。

实现拆解

1. 审查文件 `.buildkite/test-amd.yaml` 中所有测试组定义, 识别出适合从 MI250 转移到 MI300 的测试组。
2. 修改 `mirror_hardware` 字段: 将 Pipeline + Context Parallelism (4 GPUs) 等测试组的硬件列表从包含 MI250 改为仅包含 MI300 等现代硬件。
3. 删除完全冗余的测试组: 例如 PyTorch Fullgraph、PyTorch Fullgraph Smoke Test、Elastic EP Scaling Test、EPLB Execution、Multi-Modal Accuracy Eval 等, 这些测试或已不再需要, 或已在更合适的位置运行。
4. 调整后, CI 配置总行数减少 152 行 (+131/-283), 结构更简洁, 资源分配更合理。

关键文件:

- `.buildkite/test-amd.yaml` (模块 CI 配置; 类别 config; 类型 configuration): 唯一变更的文件, 包含所有 ROCm CI 测试组定义。通过删除冗余组和调整硬件分配直接实现 PR 目标。

关键符号: 未识别

评论区精华

该 PR 没有 review 评论, 直接获得批准。

- 暂无高价值评论线程

风险与影响

- 风险：主要风险在于删除的测试组可能仍有覆盖需求，但 PR 作者明确说明这些组应仅在现代硬件上运行，且部分测试已迁移。若后续发现缺失，可随时通过恢复或补充其它配置来弥补。由于只涉及 CI 编排，不影响任何运行时逻辑，回归风险极低。
- 影响：对用户无影响；对 ROCm CI 系统而言，MI250 集群负载将减轻，MI300 集群资源利用率提升；测试执行策略更加合理，有助于缩短 CI 排队时间。影响范围限定于 AMD GPU CI 流水线。
- 风险标记：测试组删除可能需验证覆盖，硬件分配变更可能影响调度

关联脉络

- PR #46417 [ROCm][CI] Increase the max wait time for server startup: 同为 ROCm CI 优化主题，调整测试配置以提高 CI 稳定性。
- PR #46141 [ROCm][CI] Query total device memory via amdsmi to avoid HIP init: 同为 ROCm CI 改进，关注 CI 性能与硬件适配。