

PR #49257 完整报告

vllm-project/vllm

[CI][AMD] Deprecate DinD for MI355 tests

合并时间: 2026-07-25 02:41

原文链接: <http://prhub.com.cn/vllm-project/vllm/pull/49257>

执行摘要

- 一句话: 为 MI355 CI 作业移除 DinD 模式
- 推荐动作: 普通维护变更, 无需精读。值得关注的是 AMD CI 基础设施的演进方向。

功能与动机

废弃 DinD 以减少 CI 基础设施的复杂性和开销, 直接在 AMD Kubernetes Pod 中运行 MI355 测试步骤, 无重复 PR 已包含此更改。

实现拆解

1. 修改 CI 配置文件: 在 `.buildkite/test-amd.yaml` 中, 为所有 `mirror_hardware`s 包含 `amdgfx950nightly` 或 `amdmi355` 的测试步骤添加 `dind: false` 参数。
2. 覆盖的测试类别: 包括 Attention Benchmarks Smoke Test、Distributed Tests (2xH100-2xMI355)、Entrypoints Integration (API Server) 及其变体、GPQA Eval (GPT-OSS)、LM Eval Qwen3-5 Models 等。
3. 不涉及代码逻辑变更: 仅修改 CI 配置, 没有影响 vLLM 运行时或其他基础设施。

关键文件:

- `.buildkite/test-amd.yaml` (模块 CI 配置; 类别 `config`; 类型 `configuration`): AMD CI 测试配置文件; 所有 MI355 测试步骤添加 `dind: false`, 废弃 DinD 模式。

关键符号: 未识别

评论区精华

无实质 review 讨论; 自动化 bot 评论因来自 fork 而跳过, `tjtanaa` 给予 Approve。

- 暂无高价值评论线程

风险与影响

- 风险: 低风险。纯 CI 配置变更, 不影响核心推理逻辑或稳定性。潜在风险是如果某些测试步骤原本依赖于 DinD 提供的隔离环境 (如网络或文件系统), 直接运行时可能失效, 但 `dind: false` 在其他 MI300 作业中已有使用 (示例中 - `label: Weight Loading Multiple GPU - Large Models` 已设置 `dind: false`), 因此经验证可行。

- 影响：仅影响 AMD MI355 CI 测试的运行方式：从 DinD 容器模式改为直接运行在 Kubernetes Pod 中。预计可减少资源开销，简化调试流程，提升 CI 效率。其他 GPU 型号（如 MI300）不受影响。
- 风险标记：暂无

关联脉络

- 暂无明显关联 PR