

PR #50311 完整报告

vllm-project/vllm

[ROCm][CI] Avoid Ray worker startup env race

合并时间: 2026-07-30 05:30

原文链接: <http://prhub.com.cn/vllm-project/vllm/pull/50311>

执行摘要

- 一句话: 修复 ROCm CI 中 Ray worker 启动环境竞争
- 推荐动作: 可快速合并。建议后续跟踪 Ray 上游的长期修复策略, 或考虑将相同约束应用于非 ROCm 的测试依赖文件以确保一致性。

功能与动机

AMD CI build 11417 中 pipeline parallelism 测试 (RayExecutorV2, 4 GPUs) 出现 Ray actor 在导入类时因 libc `getenv` 竞态而崩溃。该问题根源在 Ray 框架本身 (PR #55152, #61034, #61281), 而非 vLLM 代码。

实现拆解

1. 修改依赖约束: 在 requirements/test/rocm.in 中将 Ray 的版本从宽松的 `>=2.48.0` 收紧为精确的 `==2.55.1`, 并更新注释说明原因。
2. 同步锁定文件: 更新 requirements/test/rocm.txt 中 Ray 的解析版本从 2.54.0 改为 2.55.1, 确保实际安装的版本一致。
3. 无其他代码变更: 仅依赖版本调整, 不涉及 vLLM 运行时逻辑。

关键文件:

- requirements/test/rocm.in (模块 构建配置; 类别 test; 类型 configuration): 主变更文件: 将 Ray 版本约束从 `>=2.48.0` 改为 `==2.55.1`, 以包含上游竞态修复。
- requirements/test/rocm.txt (模块 构建配置; 类别 docs; 类型 documentation): 锁定文件同步更新: 反映 rocm.in 的版本变更, 确保实际安装版本一致。

关键符号: 未识别

评论区精华

PR 无 review 评论, 自动审阅 bot 因来自 fork 而未执行, 仅由维护者 mgoin 直接批准。

- 暂无高价值评论线程

风险与影响

- 风险：风险极低：仅更改了 ROCm CI 环境的测试依赖版本，未触及任何生产代码或非测试依赖。Ray 2.55.1 是向后兼容的补丁版本，不会引入 API 破坏性变更。但建议确认 2.55.1 与现有测试栈（如 pytest、tensorizer）的兼容性。
- 影响：影响范围局限于 ROCm CI 构建环境，特别是使用 RayExecutorV2 的 pipeline parallelism 测试。ROCM 用户和开发者将受益于更稳定的 CI，减少由于 Ray 竞态导致的随机失败。不影响其他平台或生产部署。
- 风险标记：依赖版本收紧，仅影响测试环境

关联脉络

- PR #49647 [Rubin] Enable NVLink all-reduce paths on SM107: 同属 ROCm/AMD 相关硬件支持与稳定性改进