

PR #47299 完整报告

vllm-project/vllm

[CI] Fix segfault in tracing test

合并时间: 2026-07-01 22:00

原文链接: <http://prhub.com.cn/vllm-project/vllm/pull/47299>

执行摘要

- 一句话: 修复 tracing 测试在不同平台上的段错误
- 推荐动作: 简单安全的 bugfix, 建议合并。无需深入审查, 但值得存档作为 gRPC fork 安全问题的参考。

功能与动机

CI 构建 (如 Buildkite 75740) 中非 ROCm 平台上的 tracing 测试因 gRPC 线程 fork 安全问题间歇性 segfault。之前仅在 ROCm 条件下修复, 但问题不限于 ROCm。

实现拆解

1. 移除平台条件判断: 在 tests/v1/tracing/test_tracing.py 的 test_traces 函数中, 删除 if current_platform.is_rocm() 条件, 使 VLLM_WORKER_MULTIPROC_METHOD=spawn 设置对所有平台生效。
2. 更新注释: 将原针对 ROCm 的说明改为通用描述, 指出 gRPC C-core 非 fork-safe。
3. 无其他变更: 仅修改此测试文件, 删除 1 行条件判断和 1 行缩进, 新增 1 行无缩进的 setenv 调用。

关键文件:

- tests/v1/tracing/test_tracing.py (模块 追踪; 类别 test; 类型 test-coverage): 唯一变更文件, 移除平台条件使其总是设置 spawn 模式, 从而修复跨平台 segfault。

关键符号: test_traces

关键源码片段

tests/v1/tracing/test_tracing.py

唯一变更文件, 移除平台条件使其总是设置 spawn 模式, 从而修复跨平台 segfault。

```
def test_traces(
    monkeypatch: pytest.MonkeyPatch,
    trace_service: FakeTraceService,
):
    with monkeypatch.context() as m:
        m.setenv(OTEL_EXPORTER_OTLP_TRACES_INSECURE, "true")
```

```
# 修复: gRPC C-core 在 fork 时不安全, 因此总是使用 spawn 模式
# 之前仅对 ROCm 设置, 但问题不限于 ROCm
m.setenv("VLLM_WORKER_MULTIPROC_METHOD", "spawn")
# ... 其余测试逻辑保持不变
```

评论区精华

无 review 评论或讨论, 仅 Claude bot 自动评论且未交互。hmellor 直接批准。

- 暂无高价值评论线程

风险与影响

- 风险: 风险极低。变更仅影响单测环境变量设置, 且 spawn 是 Python 多进程的跨平台安全模式, 不会影响生产代码。可能轻微增加测试启动时间 (spawn 比 fork 稍慢), 但预计可忽略。
- 影响: 影响范围限于测试稳定性。确保 tracing 测试在所有 CI 平台上可靠运行, 消除 ROCm 和 CUDA/CPU 环境之间的行为差异。对用户无影响, 对系统性能无影响。
- 风险标记: 测试环境变更新增

关联脉络

- 暂无明显关联 PR