

PR #44128 完整报告

vllm-project/vllm

[Misc] Remove dead VLLM_RPC_TIMEOUT env var and fix profiling doc that references it

合并时间: 2026-06-03 08:22

原文链接: <http://prhub.com.cn/vllm-project/vllm/pull/44128>

执行摘要

- 一句话: 移除死环境变量并修正文档
- 推荐动作: 此 PR 展示了如何安全清理死代码和同步文档, 适用于类似的遗留清理任务。

功能与动机

该变量在 V1 引擎客户端中不再被使用, 但文档仍指导用户设置, 造成误解。profiling 文档中的相关建议也已不再需要。

实现拆解

1. 删除 vllm/envs.py 中 VLLM_RPC_TIMEOUT 的类型定义 (第 98 行) 和 environment_variables 注册 (第 1018-1020 行)。
2. 修改 docs/contributing/profiling.md, 移除对 VLLM_RPC_TIMEOUT 的引用, 改为说明 engine client 会等待 trace flush 完成不会超时。
3. 从 .buildkite/performance-benchmarks/tests/ 下的 9 个 CPU 基准配置文件中移除 VLLM_RPC_TIMEOUT 环境变量设置。

关键文件:

- vllm/envs.py (模块 配置层; 类别 source; 类型 core-logic) : 核心变更: 删除死环境变量 VLLM_RPC_TIMEOUT 的类型定义和注册条目
- docs/contributing/profiling.md (模块 文档; 类别 docs; 类型 documentation) : 文档修正: 移除过时的 VLLM_RPC_TIMEOUT 设置指导
- .buildkite/performance-benchmarks/tests/latency-tests-cpu.json (模块 基准配置; 类别 test; 类型 test-coverage) : 清理基准配置中的无用环境变量 (类似配置共 9 个)
- .buildkite/performance-benchmarks/tests/serving-tests-cpu.json (模块 基准配置; 类别 test; 类型 test-coverage) : 清理基准配置中的无用环境变量 (类似配置共 9 个)

关键符号: 未识别

关键源码片段

[vllm/envs.py](#)

核心变更: 删除死环境变量 VLLM_RPC_TIMEOUT 的类型定义和注册条目

```
# 在 _EnvVars 类中移除了 VLLM_RPC_TIMEOUT 类型定义 (原第 98 行)
class _EnvVars:
    # ...
    VLLM_ALLOW_LONG_MAX_MODEL_LEN: bool = False
    # 已移除 : VLLM_RPC_TIMEOUT: int = 10000
    VLLM_HTTP_TIMEOUT_KEEP_ALIVE: int = 5 # seconds
    # ...

# 在 environment_variables 字典中移除了对应条目 (原第 1018-1020 行)
environment_variables = {
    # ...
    "VLLM_TEST_FORCE_LOAD_FORMAT": lambda: ...,
    # 已移除 :
    # "VLLM_RPC_TIMEOUT": lambda: int(os.getenv("VLLM_RPC_TIMEOUT", "10000")),
    "VLLM_HTTP_TIMEOUT_KEEP_ALIVE": lambda: ...,
    # ...
}
```

评论区精华

审核人 sfeng33 指出仍有基准配置文件设置了 VLLM_RPC_TIMEOUT，作者在最后一个提交中清理了所有 CPU 基准配置（共 9 个文件）。

- 清理所有 benchmark 配置 (other): 作者随后添加提交，从 9 个 CPU 基准配置文件中移除了该变量。

风险与影响

- 风险：风险极低：该变量已无任何消费者，删除和清理只影响无用的环境变量设置，不影响运行时行为。
- 影响：用户不再能通过设置 VLLM_RPC_TIMEOUT 影响超时行为（之前也无效）；profiling 文档更准确；CI 基准环境变量更简洁。
- 风险标记：死代码清理无功能风险

关联脉络

- PR #44279 [Refactor] Remove dead code from parser infrastructure: 同样属于清理死代码的 PR，风格和范围类似，体现持续的代码清理工作。