

PR #52672 完整报告

vllm-project/vllm

[XPU] upgrade requirements/test/xpu.txt

合并时间: 2026-08-19 10:47

原文链接: <http://prhub.com.cn/vllm-project/vllm/pull/52672>

执行摘要

- 一句话: 重编译 XPU 测试依赖, 升级 pytest 修复 caplog 失败
- 推荐动作: 值得简单阅读: 根因分析清晰展示了 pytest 日志捕获行为差异, 重新编译锁定文件对齐版本的做法合理; 但需关注传递依赖大版本跳跃带来的潜在 CI 风险, 建议合入后持续观察 XPU 测试结果。

功能与动机

PR body 明确指出: XPU CI 上 `test_arg_utils.py::test_prefix_cache_retention_interval_from_deprecated_env` 失败, CUDA CI 通过。根因是 `requirements/test/xpu.txt` 中 `pytest` 版本为 `9.0.2`, 而 `cuda.txt` 为 `9.1.0`。测试用 `caplog` 捕获 `vLLM` 的 `WARNING` 日志, 但 `vLLM` `logger` 配置了 `propagate=False` (`vllm/logger.py:67`), `pytest 9.0.2` 的 `caplog` 依赖日志向 `root logger` 传播, 导致捕获为空、断言失败; `pytest 9.1.0` 修复了此行为, 可直接捕获 `propagate=False` 的 `logger`。

实现拆解

1. 定位根因: 对比 XPU 与 CUDA 的测试依赖版本, 确认差异集中在 `pytest 9.0.2 → 9.1.0`, 并关联到 `caplog` 对 `propagate=False` 日志的捕获行为变化。
2. 重新生成锁定文件: 使用 `uv pip compile` 以 `requirements/test/xpu.in` 为输入、`requirements/xpu.txt` 为约束重新解析, 生成新的 `requirements/test/xpu.txt (+183/-169)`, 将所有测试依赖升级到与 CUDA 一致的最新版本。
3. 同步传递依赖: 升级覆盖 `absl-py`、`aiohttp`、`albumentations`、`numpy`、`openai`、`mcp` 等数十个包, 消除两平台测试环境的版本漂移。
4. 基线合并与 CI 验证: 合并 `main` 分支 (commit `00d4c7d9`) 后触发 Buildkite CI (`#84331`、`#84513`) 验证, 最终由维护者批准合入。
5. 无配套改动: 本次无源码、测试或部署配置改动, 仅依赖锁定文件变更。

关键文件:

- `requirements/test/xpu.txt` (模块 测试依赖; 类别 `config`; 类型 `dependency-update`): 唯一变更文件, 通过重新编译依赖锁定文件将 `pytest` 从 `9.0.2` 升级到 `9.1.0`, 修复 XPU CI 上 `caplog` 无法捕获 `propagate=False` 日志导致的测试失败, 同时同步更新大量传递依赖。

关键符号: 未识别

评论区精华

claude[bot] 因 PR 来自 fork 自动跳过代码审查，提示维护者可通过 @claude review 触发一次性审查；维护者 jikunshang 在触发两次 /ci run 后直接批准。关键技术论证集中在 PR body 的根因分析，即 pytest 9.0.2 与 9.1.0 在 caplog 对 propagate=False 日志捕获行为上的差异，评审中无实质技术争论。

- fork PR 自动审查被禁用 (other): 维护者 jikunshang 在 CI 触发后直接批准，未启动一次性审查。

风险与影响

- 风险：批量升级所有测试依赖可能引入 XPU 环境下的其他兼容性问题，尤其是 numpy 2.3.5、openai 3.2.0、mcp 2.0.0 等大版本跳跃；锁定文件由 uv 自动生成，若 xpu.in 或约束文件后续变化需重新编译以保证一致性；修复依赖 pytest 的具体行为，若未来 vLLM 调整 logger 配置或测试改用其他捕获方式，该修复可能不再生效。
- 影响：影响范围限于 XPU CI 测试环境，修复了 test_prefix_cache_retention_interval_from_deprecated_env 的失败；对最终用户无影响，对团队而言可降低 XPU CI 的不稳定性，并推动 XPU 与 CUDA 测试依赖版本对齐。
- 风险标记：批量依赖升级，XPU 环境特定，传递依赖大版本跳跃，仅 CI 覆盖验证

关联脉络

- PR #52797 [CI] Upgrade huggingface-hub to 1.28.0: 同样更新 requirements 测试依赖清单，属于同一依赖升级维护线，且该 PR 后 XPU 测试依赖也需要重新对齐。
- PR #52730 [XPU][CI] fix hf runner: 同为 XPU CI 稳定性的修复，目标都是在 Intel GPU 环境消除测试基础设施问题。