

PR #55799 完整报告

vllm-project/vllm

[Bugfix][CI] skip conftest for NPU compatibility test

合并时间: 2026-09-13 01:03

原文链接: <http://prhub.com.cn/vllm-project/vllm/pull/55799>

执行摘要

- 一句话: NPU CI 测试跳过 conftest 并改用 pip 修复依赖缺失
- 推荐动作: 值得快速浏览: 它展示了一个 CI 故障的两个典型根因 (conftest 传递依赖、uv 与系统 Python 解耦) 及最小化修复路径。可重点关注 `--noconftest` 副作用的评估方式, 以及在硬件 CI 中从 uv 回退到系统 pip 的取舍依据。无需深入精读。

功能与动机

PR body 明确说明失败原因: The current NPU CI test fails with `ModuleNotFoundError: No module named 'huggingface_hub'`。vllm-interface 测试用例不依赖该组件, 因此通过 `--noconftest` 跳过 `conftest.py` 加载是更小代价的修复。后续 CI 日志又暴露 `Dockerfile:15` 的 uv 报错 `No system Python installation found`, 需同步将构建依赖从 uv 迁移到系统 pip。

实现拆解

本 PR 仅修改 `.buildkite/scripts/hardware_ci/run-npu-test.sh`, 按以下步骤完成修复:

1. 定位两个失败根因: PR body 指出 NPU CI 报 `ModuleNotFoundError: No module named 'huggingface_hub'`, 推测是仓库 `conftest.py` 导入链拉入该模块, 而 `vllm-interface` 测试并不需要它; 后续 CI 日志又暴露第二个根因——`Dockerfile` 构建阶段 `uv pip install` 报 `No system Python installation found` (NPU 基础镜像未预装系统 Python, uv 无法定位解释器)。
2. Docker 构建阶段从 uv 迁移到系统 pip: 删除 uv 安装脚本与 `UV_SYSTEM_PYTHON`、`UV_INDEX_URL`、`UV_EXTRA_INDEX_URL` 等环境变量, 改用 `pip config set` 配置内部 PyPI 源和 `trusted-host`; 同时把带 `cache mount` 的 `uv pip install pytest` 与 `modelscope<1.38` 替换为 `pip install pytest>=6.0`, `cache` 挂载路径也相应从 `/root/.cache/uv` 调整为 `/root/.cache/pip`。
3. 测试运行命令绕过 conftest: 将 `pytest -v -s tests/e2e/vllm_interface/test_vllm_pr_interface_compatibility.py` 改为 `pytest --noconftest -v -s ...`, 跳过仓库 `conftest.py` 加载, 从源头避免 `huggingface_hub` 导入。
4. 配套与验证: 随依赖裁剪同步调整了 Docker cache 挂载路径; 作者完成本地测试, 维护者触发多轮 CI 后批准合并。此变更无源码、schema 或运行时配置联动。

关键文件:

- `.buildkite/scripts/hardware_ci/run-npu-test.sh` (模块 测试脚本; 类别 `infra`; 类型 `infrastructure`) : 本 PR 唯一变更文件, 承载两处根因修复: Docker 构建阶段从 `uv` 迁移到系统 `pip` 并裁剪 `modelscope` 依赖, 测试命令追加 `--noconftest` 跳过 `conftest.py` 加载。

关键符号: 未识别

关键源码片段

`.buildkite/scripts/hardware_ci/run-npu-test.sh`

本 PR 唯一变更文件, 承载两处根因修复: Docker 构建阶段从 `uv` 迁移到系统 `pip` 并裁剪 `modelscope` 依赖, 测试命令追加 `--noconftest` 跳过 `conftest.py` 加载。

```
# NPU 兼容性测试镜像构建与运行片段 (run-npu-test.sh)
# 原实现依赖 uv, 但 NPU 基础镜像未预装系统 Python。
# Dockerfile 构建阶段执行 uv pip install 时报 No system Python installation found。
# 因此改回系统 pip, 并显式配置内部 PyPI 源与 trusted-host

RUN pip config set global.index-url \
    http://cache-service-vllm.nginx-pypi-cache.svc.cluster.local:${PYPI_CACHE_PORT}/pypi/
    simple && \
    pip config set global.trusted-host \
    cache-service-vllm.nginx-pypi-cache.svc.cluster.local

# vllm-interface 测试只需要 pytest, 顺带移除原 uv 安装的 modelscope<1.38 依赖
RUN --mount=type=cache,target=/root/.cache/pip \
    pip install pytest>=6.0

# 运行测试: --noconftest 跳过仓库 conftest.py。
# 避免其传递导入 huggingface_hub。该测试用例本身不依赖 conftest 提供的 fixture
pytest --noconftest -v -s tests/e2e/vllm_interface/test_vllm_pr_interface_compatibility.py
```

评论区精华

核心讨论围绕 CI 失败日志展开: ZJY0516 作为维护者两次触发 CI (Buildkite #87846、#87954), 并在评论中贴出 Dockerfile:15 的构建失败日志, 定位到 `uv pip install` 报 `No system Python installation found`、`exit code 2`, 最终推动将 `uv` 替换为系统 `pip`。claude[bot] 因 fork PR 自动 review 被禁用, 未产生有效评审。ZJY0516 最终以 `APPROVED` 状态批准合并。

- `conftest.py` 引入 `huggingface_hub` 导致 NPU CI 失败 (testing): 在 `pytest` 命令中追加 `--noconftest`, 跳过仓库 `conftest.py` 的加载
- Docker 构建阶段 `uv` 找不到系统 Python (testing): 将 `uv` 安装与配置替换为系统 `pip` `config` 与 `pip install pytest`, 并移除 `modelscope<1.38` 依赖

风险与影响

- 风险:

1. `--noconftest` 是全局行为：它会跳过仓库及测试目录下所有 `conftest.py`，而不仅是触发 `huggingface_hub` 的那一个。若 `vllm-interface` 测试未来依赖 `conftest` 提供的 `fixture`，可能出现测试环境不完整导致的假阳性，当前用例声明不依赖。
2. `uv` 到 `pip` 的迁移改变了 Docker 构建缓存键与缓存挂载路径（`/root/.cache/uv` 变为 `/root/.cache/pip`），首次重新构建需要重新拉包，且 `pip` 与 `uv` 的依赖解析行为不同，后续新增依赖时需留意。
3. `modelscope<1.38` 依赖被移除：PR 声明测试不需要，但若未来用例扩展为真实模型下载，会缺失该依赖。
4. 内部 PyPI 镜像地址为集群内 `service`，脚本只能在 NPU CI 内网环境复现，本地调试需额外配置。
 - 影响：影响范围仅限 NPU 硬件 CI 的 `vllm-interface` 兼容性测试（`tests/e2e/vllm_interface/test_vllm_pr_interface_compatibility.py`）。对最终用户无任何影响，对维护者而言解除了 NPU CI 持续失败导致的阻塞，降低人工排查成本。变更不涉及运行时代码，风险隔离在单条 CI 脚本内，不影响其他硬件平台。
 - 风险标记：跳过 `conftest` 加载，构建依赖变更，移除 `modelscope` 依赖，仅 NPU 平台生效

关联脉络

- PR #56463 [XPU][CI] skip DeepSeek-V4.1-Flash in `test_tensor_schema.py`: 同为硬件平台 CI 中跳过必失败用例的修复模式，与本 PR 的 `--noconftest` 思路一致，反映硬件 CI 的持续收敛
- PR #55171 [XPU][CI] Remove `pip install` dependency in test yaml files: 同样收敛硬件平台 CI 的依赖安装方式，与本 PR 的 `pip` 依赖调整同属 CI 环境治理
- PR #56555 [XPU][CI] Skip ROCm test on non-ROCM platforms: 同类跨硬件平台 CI 修复，体现 vLLM 对非 NVIDIA 平台 CI 稳定性的持续投入