

# PR #43703 完整报告

vllm-project/vllm

[CI][ROCm] Don't skip MoRI-IO Connector tests

合并时间: 2026-05-29 17:06

原文链接: <http://prhub.com.cn/vllm-project/vllm/pull/43703>

## 执行摘要

- 一句话: 修复 ROCm CI 中 MoRI 单元测试被错误跳过的问题
- 推荐动作: 可供快速合并, 无重大问题。建议 CI 相关人员了解变更背景, 并关注后续 e2e 测试是否按 PR 描述使用 RDMA 后端。

## 功能与动机

PR 描述指出: MoRIIO 连接器单元测试在 AMD CI runner 上被跳过, 即使主机有 RNIC, 因为 `ibv_devinfo` 命令在 vLLM ROCm 镜像中未安装。这些单元测试仅验证连接器功能, 而非 RDMA 传输, 因此可以使用任意 MoRI 后端。

## 实现拆解

1. 删除 `_rdma_available()` 函数及相关导入: 移除 `import subprocess` 和依赖 `ibv_devinfo` 的 `_rdma_available()` 函数及模块级变量 `rdma_available`。
2. 移除两个 `@pytest.mark.skipif` 装饰器: 从 `test_register_kv_caches` 和 `test_moriio_handshake_returns_metadata` 中移除 RDMA 可用性条件跳过。
3. 在 `create_vllm_config` 中添加 `backend: xgmi` 配置: 将 `kv_connector_extra_config` 从 `{"read_mode": read_mode}` 改为 `{"read_mode": read_mode, "backend": "xgmi"}`, 并添加注释说明原因: xGMI 后端不依赖 RNIC, 适合 CI 环境。
4. 更新模块级测试跳过条件: 保留 `pytestmark` 中 `current_platform.is_rocm()` and `mori_available` 的判断。

关键文件:

- `tests/v1/kv_connector/unit/test_moriio_connector.py` (模块 KV 连接器; 类别 `test`; 类型 `test-coverage`; 符号 `_rdma_available`): 唯一变更文件, 修改了 RDMA 可用性检查逻辑和连接器后端配置, 以解决 CI 跳过问题。

关键符号: `_rdma_available`

## 关键源码片段

`tests/v1/kv_connector/unit/test_moriio_connector.py`

唯一变更文件, 修改了 RDMA 可用性检查逻辑和连接器后端配置, 以解决 CI 跳过问题。

```
# tests/v1/kv_connector/unit/test_moriio_connector.py
```

```

import importlib.util
import uuid
from unittest.mock import MagicMock, patch

import pytest
import torch
import zmq

# ... 其他 import ...

aiter_available = importlib.util.find_spec("aiter") is not None
mori_available = importlib.util.find_spec("mori") is not None

# 模块级别跳过条件：仅在 ROCm 且安装了 mori 包时执行测试
pytestmark = pytest.mark.skipif(
    not (current_platform.is_rocm() and mori_available),
    reason="MoRIIOs are only available on ROCm with aiter package installed",
)

def create_vllm_config(...):
    # ... 其他配置 ...
    # 注释说明：这些单元测试仅验证连接器 setup，不涉及真实 RDMA 传输
    # (MoRI 包装器被 mock)，因此可以使用任意后端而不影响测试有效性。
    # 使用 xGMI 以避免在 CI 中要求 RNIC。
    kv_transfer_config = KVTransferConfig(
        kv_connector="MoRIIOConnector",
        kv_role=role,
        enable_permute_local_kv=enable_permute_local_kv,
        kv_connector_extra_config={"read_mode": read_mode, "backend": "xgmi"},
    )
    return VllmConfig(...)

```

## 评论区精华

PR 没有 review 评论。两位 reviewer (AndreasKaratzas 和 tjtanaa) 均批准，无讨论。

- 暂无高价值评论线程

## 风险与影响

- 风险：风险极低。变更仅涉及测试配置和跳过条件，不触及任何生产代码或核心逻辑。唯一风险是：若未来 MoRIIO 连接器的单元测试依赖于 RDMA 后端特定的行为，使用 xGMI 可能导致测试与真实环境不完全一致。但 PR 描述和注释已明确说明当前测试仅验证连接器管理功能，不涉及 RDMA 传输，且未来 e2e 测试将使用 RDMA 后端。
- 影响：影响范围：仅限 ROCm CI 上的 MoRIIO 连接器单元测试。影响程度低：
  - CI 可靠性提升：确保 5 个被测用例在具备 RNIC 的 ROCm CI runner 上不再被跳过。
  - 无用户影响：不变更生产运行时行为。
  - 无性能影响：测试不涉及性能关键路径。

- 无兼容性问题：xGMI 后端在 MoRPIO 库中已存在。
- 风险标记：仅测试变更，低风险

## 关联脉络

- PR #43797 [kv\_offload] Skip decode-phase blocks in CPU offload: 同为 KV 连接器相关变更，修改了同一模块目录下的文件。