

# PR #41076 完整报告

vllm-project/vllm

[CI][AMD][BugFix] Update request URL in test\_moriiio\_connector to match vllm-router compatibility changes

合并时间: 2026-04-29 02:14

原文链接: <http://prhub.com.cn/vllm-project/vllm/pull/41076>

## 执行摘要

- 一句话: 修复 moriio 连接器测试的请求 URL 格式
- 推荐动作: 此 PR 为纯粹的测试兼容性修复, 变更清晰直接, 值得快速合入以确保 CI 健康。可精读的亮点在于通过调整 request\_id 构造时机, 避免了调度器内部缓存不一致的潜在问题。

## 功能与动机

PR #39565 修改了 moriio KV 连接器, 使其兼容 vllm-router, 改变了请求 URL 格式以包含 prefill、decode 和 uuid 值。但 test\_moriiio\_connector 中的测试未更新, 生成的请求未包含 URL 变更以保持兼容性。此 PR 更新测试以生成与新模式匹配的请求, 避免回归。

## 实现拆解

1. 新增 uuid 导入: 在 tests/v1/kv\_connector/unit/test\_moriiio\_connector.py 文件顶部添加 import uuid。
2. 修改 \_setup\_kv\_transfer\_request 辅助函数: 在更新 kv\_transfer\_params 后, 构造 ZMQ 地址字符串 zmq\_addr, 生成随机 UUID, 并将 request.request\_id 设置为 f"\_\_prefill\_addr\_{zmq\_addr}\_\_decode\_addr\_{zmq\_addr}\_{fake\_uuid}"。
3. 调整三个测试用例的执行顺序: 将 \_setup\_kv\_transfer\_request(request) 的调用从 scheduler.add\_request(request) 之后移到之前, 确保 request\_id 在请求加入调度器之前就已包含正确的格式, 避免调度器内部使用旧 ID 导致不一致。受影响的测试包括 test\_write\_mode\_saves\_local\_block\_ids、test\_write\_mode\_with\_chunked\_prefill\_saves\_local\_block\_ids 和 test\_read\_mode\_loads\_remote\_block\_ids。

关键文件:

- tests/v1/kv\_connector/unit/test\_moriiio\_connector.py (模块 KV 连接器; 类别 test; 类型 test-coverage; 符号 \_setup\_kv\_transfer\_request, test\_write\_mode\_saves\_local\_block\_ids, test\_write\_mode\_with\_chunked\_prefill\_saves\_local\_block\_ids, test\_read\_mode\_loads\_remote\_block\_ids): 所有变更集中于此文件, 包括新增 uuid 导入、修改 \_setup\_kv\_transfer\_request 以嵌入 ZMQ 地址和 UUID 到 request\_id, 以及调整三个测试用例的调用顺序, 确保测试请求格式与 vllm-router 兼容。

关键符号: `_setup_kv_transfer_request`, `test_write_mode_saves_local_block_ids`,  
`test_write_mode_with_chunked_prefill_saves_local_block_ids`,  
`test_read_mode_loads_remote_block_ids`

## 关键源码片段

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

所有变更集中于此文件, 包括新增 `uuid` 导入、修改 `_setup_kv_transfer_request` 以嵌入 ZMQ 地址和 UUID 到 `request_id`, 以及调整三个测试用例的调用顺序, 确保测试请求格式与 `vllm-router` 兼容。

```
def _setup_kv_transfer_request(
    request, remote_host="127.0.0.1", fake_port=4789, fake_transfer_id="0"
):
    """Setup KV transfer parameters for a request."""
    # 更新 kv_transfer_params, 填充测试用的远程地址和端口等元数据
    request.kv_transfer_params.update(
        {
            "transfer_id": fake_transfer_id,
            "remote_notify_port": fake_port,
            "remote_block_ids": None,
            "remote_host": remote_host,
            "remote_port": fake_port,
            "remote_handshake_port": fake_port,
            "remote_engine_id": "test_engine",
        }
    )
    # 构造 ZMQ 地址字符串, 格式与 vllm-router 兼容
    zmq_addr = f"host:{remote_host},handshake:{fake_port},notify:{fake_port}"
    fake_uuid = uuid.uuid4().hex # 生成随机 UUID
    # 将 ZMQ 地址和 UUID 嵌入 request_id, 以匹配生产代码中的 URL 格式
    request.request_id = (
        f"__prefill_addr_{zmq_addr}__decode_addr_{zmq_addr}_{fake_uuid}"
    )
    return request
```

## 评论区精华

- 暂无高价值评论线程

## 风险与影响

- 风险: 风险极低。变更仅涉及测试文件, 未修改生产代码。但需注意新 `request_id` 格式依赖于 `_setup_kv_transfer_request` 的调用时机, 若未来测试用例复用此辅助函数但未遵循调用顺序, 可能出现请求 ID 不包含预期格式的问题。
- 影响: 影响范围仅限于 `moriio KV` 连接器的单元测试, 确保测试请求 URL 格式与生产代码兼容, 避免 CI 因格式不匹配而失败。对用户和系统无直接影响。

- 风险标记：仅测试变更

## 关联脉络

- PR #39565 Moriiovllm router compatible: 本次测试修复的直接原因：PR #39565 修改了 moriio KV 连接器以兼容 vllm-router，改变了请求 URL 格式，但未更新对应测试。
- PR #37243 Moriiovllm connector: Issue 评论中提及该 PR 修复了一个关键回归（链接指向此 PR），可见 #39565 之前的变更引入了回归。