

PR #47222 完整报告

vllm-project/vllm

[Hardware][AMD][CI] Toggle test coredumps on ROCm debug agent

合并时间: 2026-07-01 12:30

原文链接: <http://prhub.com.cn/vllm-project/vllm/pull/47222>

执行摘要

- 一句话: AMD CI 测试容器 GPU coredump 开关优化
- 推荐动作: 建议合并。该 PR 是针对 AMD CI 稳定性问题的定向修复, 逻辑清晰, 风险可控。

功能与动机

AMD CI 测试中, 内存访问故障后 GPU coredump handler 因 ROCm 运行时 pipe handler 二进制缺失 (execvp failed) 而无法正常退出, 导致测试挂起。关联 rocm-systems issue#6206。本 PR 旨在消除此干扰, 提升 CI 稳定性。

实现拆解

1. 在 .buildkite/scripts/hardware_ci/run-amd-test.sh 中新增 coredump 开关逻辑: 通过 ulimit -H -c 获取硬限制, 若为 unlimited 则转为 -1 以便 docker 接受; 默认设置 --ulimit core=0:\$ulimit_core_hard 禁用 coredump; 若启动命令中包含 "ROCm debug agent enabled", 则改为设置 -e HSA_COREDUMP_PATTERN="/tmp/gpucore.%p" 环境变量以绕开 issue#6206。
2. 在 tests/distributed/test_weight_transfer.py 中, 将 torch.accelerator.set_device(device) 替换为 torch.accelerator.set_device_index(device), 以兼容 AMD GPU 的 API 要求, 使该测试可在 AMD CI 中通过。
3. 该 PR 包含 3 次提交, 分别对应 coredump 开关逻辑、unlimited 处理修复、以及 set_device_index 兼容性修改。

关键文件:

- .buildkite/scripts/hardware_ci/run-amd-test.sh (模块 CI 脚本; 类别 infra; 类型 infrastructure): 核心变更文件: 新增 coredump 开关逻辑, 根据 ROCm debug agent 启用状态动态设置 docker run --ulimit 或环境变量, 控制 GPU coredump 行为。
- tests/distributed/test_weight_transfer.py (模块 权重传输; 类别 test; 类型 test-coverage; 符号 _set_ray_assigned_device): 附带修改: 将 torch.accelerator.set_device 替换为 set_device_index, 以兼容 AMD GPU, 使该测试在 AMD CI 中通过。

关键符号: 未识别

关键源码片段

[.buildkite/scripts/hardware_ci/run-amd-test.sh](#)

核心变更文件：新增 coredump 开关逻辑，根据 ROCm debug agent 启用状态动态设置 docker run --ulimit 或环境变量，控制 GPU coredump 行为。

```
# 获取 core 文件大小的硬限制
ulimit_core_hard=$(ulimit -H -c)
# docker run 无法传递 "unlimited" 给 --ulimit
if [[ "$ulimit_core_hard" == "unlimited" ]]; then
    ulimit_core_hard="-1"
fi

# 默认禁用 coredump（硬限制保留），除非启用了 ROCm debug agent
coredump_flags="--ulimit core=0:$ulimit_core_hard"
if [[ "$commands" == *"ROcm debug agent enabled"* ]]; then
    # 绕开 https://github.com/rocm/rocm-systems/issues/6206
    # 直接指定文件路径而非使用 kernel core_pattern 的 pipe handler
    coredump_flags='-e HSA_COREDUMP_PATTERN="/tmp/gpucore.%p"'
else
    echo "ROcm debug agent not enabled, coredumps are disabled in the test container."
fi

docker run \
    --device /dev/kfd $BUILDKITE_AGENT_META_DATA_RENDER_DEVICES \
    $RDMA_FLAGS \
    --network=host \
    --shm-size=16gb \
    --group-add "$render_gid" \
    --rm \
    $coredump_flags \ # 插入 coredump 开关标志
    -e HF_TOKEN \
    ...
```

[tests/distributed/test_weight_transfer.py](#)

附带修改：将 torch.accelerator.set_device 替换为 set_device_index，以兼容 AMD GPU，使该测试在 AMD CI 中通过。

```
def _set_ray_assigned_device() -> torch.device:
    device = _get_ray_assigned_device()
    # 使用 set_device_index 替代 set_device 以兼容 AMD GPU
    # 该 API 在 PyTorch 中更通用，不会在非 NVIDIA 平台上触发错误
    torch.accelerator.set_device_index(device)
    return device
```

评论区精华

无实质性 review 讨论。作者自评指出 test_weight_transfer.py 的修改与主目的正交，但在 AMD CI 中需要此变更以使 [mi300_2: Distributed Tests](#) 测试组通过。AndreasKaratzas 批准了该 PR。

- 暂无高价值评论线程

风险与影响

- 风险：低风险。变更仅限于 AMD CI 测试容器启动脚本和一处测试文件，coredump 开关逻辑仅在 ROCm 平台生效，默认禁用 coredump 不会影响正常测试功能，仅消除因 coredump handler 失败导致的挂起。test_weight_transfer.py 的 API 替换是 torch 官方推荐的兼容用法，不改变行为。
- 影响：影响范围局限于 AMD CI 测试环境，提升稳定性，避免因 GPU coredump 问题导致的测试挂起和超时。对其他平台无影响。
- 风险标记：暂无

关联脉络

- PR #383 [Infra] Add VLLM_CI_ENABLE_ROCM_DEBUG_AGENT env var: 该 PR 是此 CI 修复的前提，定义了 VLLM_CI_ENABLE_ROCM_DEBUG_AGENT 环境变量，本 PR 的 coredump 开关逻辑依赖于它。