PR #2671 完整报告

radixark/miles

[AMD] Enable four CI tests on ROCm

合并时间：2026-08-21 05:05

原文链接：http://prhub.com.cn/radixark/miles/pull/2671

执行摘要

- 一句话：启用四个 AMD CI 测试并修复相关配置
- 推荐动作：此 PR 值得关注其 CI 基础设施配置方式，特别是如何通过调整 EP 解决并行度不匹配问题，以及如何在 Docker 和 CI 中统一添加依赖。对于负责 CI 维护的工程师，可从中学习到测试启用的标准流程。

功能与动机

团队希望扩展 AMD CI 覆盖范围，使更多测试能在 ROCm 平台上运行。PR 描述明确提到要启用四个测试，并解决它们之前失败的问题：MTP 因 world_size (4) is not divisible by expert_tensor_model_pipeline_parallel size (8) 失败，Inkling 因缺少 numba 模块失败，而 offload 和 FSDP 测试则通过上游修复直接通过。

实现拆解

1. 调整 MTP 测试的专家并行度：在 tests/e2e/megatron/test_qwen3_5_35B_A3B_mtp/test_amd_mtp1_spec_v2_r3.py 中，将 ep_size 从 4 改为 2，并简化了关于并行度配置的注释，确保 PP2 * EP2 * expert-TP1 等于 world size 4。
2. 移除测试的禁用标记：从四个测试文件中移除 disabled 参数（原本标记为 FIXME），使它们能够参与 ROCm CI。
3. 在 CI 工作流中安装 numba：在 .github/workflows/_run-ci-rocm.yml 的两个安装路径中，将 pip install 命令中加入 "numba==0.65.1"，以解决 Inkling 处理器对 SGLang 依赖的缺失。
4. 在 Docker 镜像中安装 numba：在 docker/Dockerfile.rocm 中，在 SGLang 安装后添加 RUN pip install "numba==0.65.1"，确保镜像中也有该依赖。
5. 测试与部署配套：除上述修改外，没有新增或修改测试用例本身，仅调整了 CI 注册配置。

关键文件：

- tests/e2e/megatron/test_qwen3_5_35B_A3B_mtp/test_amd_mtp1_spec_v2_r3.py（模块 测试；类别 test；类型 test-coverage）：核心修改文件，调整 EP 以解决并行度不匹配问题，并移除禁用标记
- tests/e2e/megatron/model_scripts/test_inkling_small_4layer_ci.py（模块 测试；类别 test；类型 test-coverage）：移除禁用标记，重新启用测试，依赖 numba 支持

- tests/e2e/fsdp/test_qwen3_4B_fsdp_hybrid_shard_r2s2.py（模块 测试；类别 test；类型 test-coverage）：移除禁用标记，重新启用 FSDP 测试
- tests/e2e/megatron/test_qwen3_4B_offload_disk_stream.py（模块 测试；类别 test；类型 test-coverage）：移除禁用标记，重新启用磁盘流式卸载测试
- .github/workflows/_run-ci-rocm.yml（模块 CI 配置；类别 infra；类型 infrastructure）：在 CI 安装命令中加入 numba 依赖，解决 Inkling 导入失败
- docker/Dockerfile.rocm（模块 Docker；类别 infra；类型 infrastructure）：在 ROCm Docker 镜像中安装 numba，确保测试环境一致性

关键符号：未识别

关键源码片段

tests/e2e/megatron/test_qwen3_5_35B_A3B_mtp/test_amd_mtp1_spec_v2_r3.py

核心修改文件，调整 EP 以解决并行度不匹配问题，并移除禁用标记

```
"""AMD 4-GPU variant of test_mtp1_spec_v2_r3.py.

Qwen3.5-35B-A3B: 1 MTP layer + speculative-v2 + R3, on 4 GPUs.

Standalone rather than an IS_HIP branch in the original: the MI300X fleet is
split into two 4-GPU runners, so the 8-GPU CUDA case cannot run there as
written, and keeping the variant separate means neither side's parallelism
constrains the other.

Compared with the CUDA case, the world size drops from 8 to 4, so training EP
drops from 4 to 2. This keeps PP2 * EP2 * expert-TP1 equal to world size 4.
"""

import os

from tests.ci.ci_register import register_rocm_ci
from tests.ci.metric_history import register_ci_gate
from tests.e2e.megatron.test_qwen3_5_35B_A3B_mtp._common import CaseConfig, execute,
prepare

# 注册 ROCm CI，移除 disabled 标记，使测试可以运行
register_rocm_ci(
    est_time=1600,
    suite="stage-c-4-gpu-mi350",
    labels=["megatron", "qwen35", "amd"],
)

register_ci_gate(metric_key="train/grad_norm")
register_ci_gate(metric_key="train/ppo_kl")
register_ci_gate(metric_key="train/train_rollout_logprob_abs_diff")
register_ci_gate(metric_key="train/train_rollout_kl")
```

```
register_ci_gate(metric_key="rollout/raw_reward")

CASE = CaseConfig(
    num_gpus_per_node=4,
    cp_size=1,
    pp_size=2,
    tp_size=2,
    ep_size=2, # 从 4 调整为 2，使 PP2 * EP2 * expert-TP1 = 4，匹配 world size
    rollout_num_gpus_per_engine=4,
    sglang_ep_size=4,
    enable_mtp_training=True,
    use_r3=True,
    extra_args=("--moe-token-dispatcher-type alltoall " "--sglang-disable-shared-experts-fusion "),
    # miles 在训练端没有 VLM/vision 实现，因此视觉权重永远不会同步；
    # 将其排除在权重相等性检查之外。
    check_weight_update_skip_list=("visual",),
)


if __name__ == "__main__":
    for proxy_var in ("http_proxy", "https_proxy", "HTTP_PROXY", "HTTPS_PROXY"):
        os.environ.pop(proxy_var, None)
    prepare(CASE)
    execute(CASE, wandb_file=__file__)
```

评论区精华

claude[bot] 的自动审查确认了修改的正确性：EP 调整后 expert-tensor-parallel-size(1) * ep_size(2) * pp_size(2) = 4，与 4 GPU world size 匹配；同时确认 numba 的 pin 同时在 Dockerfile 和 CI 工作流的两个路径中一致添加。guapisolo 手动审查后批准了 PR。

- EP 并行度调整的正确性 (correctness): 确认修复正确。
- numba 依赖的添加 (infrastructure): 确认添加一致，无问题。

风险与影响

- 风险：
  1. 回归风险：这些测试此前在 ROCm 平台上被禁用，现重新启用，如果运行环境（如依赖版本、硬件配置）有细微差异，可能引入新的失败。尽管在 MI355X 上已通过，但其他 ROCm runner 上可能仍存在问题。
  2. 依赖锁定风险：在 CI 和 Docker 镜像中硬编码 numba==0.65.1，可能会与未来其他依赖的更新发生冲突，但版本锁定本身也保证了可重复性。
  3. 并行度调整的影响：test_amd_mtp1_spec_v2_r3 中 EP 从 4 调整为 2，可能改变模型训练的行为（如专家并行计算分布），但这是必要的修复，且已通过测试验证。 - 影响：影响范围：仅影响 CI 配置和测试注册，不涉及任何生产代码或用户功能。影响程度：中低。对开发团队而言，扩大了 AMD CI 的测试覆盖，提升了 ROCm 平台上的代码质量保障；对正在使用 ROCm 的开发者，增强了测试的可靠性和可观察性。

- 风险标记：回归风险（重新启用测试），依赖版本锁定

关联脉络

- 暂无明显关联 PR