

# PR #6672 完整报告

verl-project/verl

[ci] chore: add two-node Ascend RayJob E2E workflow

合并时间: 2026-06-16 15:23

原文链接: <http://prhub.com.cn/verl-project/verl/pull/6672>

## 执行摘要

- 一句话: 新增双节点 Ascend NPU RayJob E2E CI 工作流
- 推荐动作: 建议 CI 维护者仔细检查工作流依赖的基础设施是否稳定, 并考虑采纳 review 中关于 shell 脚本健壮性的建议。对于希望了解 verl 如何集成 RayJob 进行多节点测试的开发者, 此 PR 提供了一份可参考的实现。

## 功能与动机

该项目缺少 Ascend NPU 上的多节点 E2E 测试, 此 PR 旨在填补该空白, 通过在 CI 中定期运行双节点 PPO 训练来提前发现多节点环境下的兼容性故障。PR body 中提及 "Search for similar PRs" 表明该工作流是经过调研后新增的。

## 实现拆解

1. 新增 CI 工作流 (.github/workflows/e2e\_ascend\_multinode.yml): 定义 workflow 触发的条件 (main 分支 push/PR, 限定特定路径), 设置并发和权限。主要 job ascend-rayjob-e2e 在专用 runner 上运行, 使用容器执行。步骤包括: 检出代码、准备共享缓存、创建 submit\_and\_wait\_rayjob.sh 辅助脚本 (用于通过 REST API 提交 RayJob 并等待其完成, 同时流式输出日志)、最后执行测试脚本并检查最终状态。
2. 新增测试脚本 (tests/special\_npu/run\_qwen3\_06b\_ppo\_multinode.sh): 该脚本封装了多节点 PPO 训练命令, 使用 verl 的 main\_ppo 入口, 配置为 Qwen3-0.6B 模型、2 节点各 4 GPU、训练 1 步。脚本使用环境变量允许外部覆盖模型路径等参数, 并利用 skip.rollout 功能跳过 rollout 阶段以加速测试 (仅验证训练流程完整性)。脚本中包含了必要的 FSDP 配置、Ulysses 序列并行等 Ascend 常用设置。
3. 集成方式: 工作流在 runner 上通过 kubectl 或类似工具与 RayJob 提交器交互 (由辅助脚本实现), 将测试脚本作为 RayJob 的任务提交, 等待完成后根据失败标志决定 CI 结论。工作流会流式输出 RayJob 的日志到 GitHub Actions 日志, 方便调试。

关键文件:

- .github/workflows/e2e\_ascend\_multinode.yml (模块 CI 配置; 类别 infra; 类型 infrastructure): 核心 CI 工作流文件, 定义了触发条件、执行环境和 RayJob 提交流程, 是本次 PR 的主要产出。
- tests/special\_npu/run\_qwen3\_06b\_ppo\_multinode.sh (模块 测试脚本; 类别 test; 类型 test-coverage): 测试脚本, 实际运行多节点 PPO 训练命令, 是工作流的核心负载。

关键符号：未识别

## 关键源码片段

`tests/special_npu/run_qwen3_06b_ppo_multinode.sh`

测试脚本，实际运行多节点 PPO 训练命令，是工作流的核心负载。

```
#!/usr/bin/env bash
# set -x # 开启命令回显（调试用）
# 注意：未使用 set -euo pipefail, review 建议添加

MODEL_ID=${MODEL_ID:-Qwen/Qwen3-0.6B}
MODEL_PATH=${MODEL_PATH:-${HOME}/.cache/models/${MODEL_ID}}

SKIP_ENABLE=True
SKIP_DUMP_DIR=${SKIP_DUMP_DIR:-${HOME}/data/rollout_dump_sync}
SKIP_STEPS='[1]' # 方括号在未引用的变量中可能被 shell 视为 glob 模式
SKIP_ACTION=cache

python3 -m verl.trainer.main_ppo \
    algorithm.adv_estimator=gae \
    data.train_files="${HOME}/data/gsm8k/train.parquet" \
    data.val_files="${HOME}/data/gsm8k/test.parquet" \
    data.train_batch_size=16 \
    data.max_prompt_length=512 \
    data.max_response_length=128 \
    data.shuffle=False \
    actor_rollout_ref.model.path="${MODEL_PATH}" \
    actor_rollout_ref.model.use_remove_padding=True \
    actor_rollout_ref.model.enable_gradient_checkpointing=True \
    actor_rollout_ref.actor.optim.lr=1e-6 \
    actor_rollout_ref.actor.ppo_mini_batch_size=8 \
    actor_rollout_ref.actor.ppo_micro_batch_size_per_gpu=1 \
    actor_rollout_ref.actor.fsdps_config.param_offload=True \
    actor_rollout_ref.actor.fsdps_config.optimizer_offload=True \
    actor_rollout_ref.actor.use_kl_loss=False \
    actor_rollout_ref.actor.ulysses_sequence_parallel_size=2 \
    actor_rollout_ref.actor.use_dynamic_bsz=True \
    actor_rollout_ref.actor.use_torch_compile=False \
    actor_rollout_ref.rollout.log_prob_micro_batch_size_per_gpu=1 \
    actor_rollout_ref.rollout.tensor_model_parallel_size=1 \
    actor_rollout_ref.rollout.name=vllm \
    actor_rollout_ref.rollout.gpu_memory_utilization=0.8 \
    actor_rollout_ref.rollout.log_prob_use_dynamic_bsz=True \
    actor_rollout_ref.rollout.enable_chunked_prefill=True \
    actor_rollout_ref.rollout.enforce_eager=False \
    +actor_rollout_ref.rollout.engine_kwargs.vllm.compilation_config.cudagraph_mode="FULL_
    AND_PIECEWISE" \
    critic.optim.lr=1e-5 \
```

```
critic.model.use_remove_padding=True \
critic.model.path="${MODEL_PATH}" \
critic.model.enable_gradient_checkpointing=True \
critic.ppo_micro_batch_size_per_gpu=1 \
critic.ulysses_sequence_parallel_size=2 \
critic.fsdparam_offload=True \
critic.fsdparam_optimizer_offload=True \
critic.use_dynamic_bsz=True \
trainer.critic_warmup=0 \
trainer.logger='["console"]' \
trainer.project_name='verl_ppo_example_gsm8k_qwen3' \
trainer.experiment_name='qwen3_06b_fsdparam_multinode_2x4' \
trainer.n_gpus_per_node=4 \
trainer.nnodes=2 \
trainer.save_freq=-1 \
trainer.test_freq=-1 \
trainer.total_epochs=1 \
trainer.total_training_steps=1 \
skip.rollout.enable=${SKIP_ENABLE} \
skip.rollout.dump_dir=${SKIP_DUMP_DIR} \
skip.rollout.steps=${SKIP_STEPS} \
skip.rollout.action=${SKIP_ACTION} \
"$@"
```

## 评论区精华

- Shell 脚本健壮性：gemini-code-assist[bot] 建议在 shell 脚本中启用 `set -x` 和 `set -e` 以尽早捕获错误，并对所有变量引用添加双引号，特别是 `SKIP_STEPS='[1]'` 可能因通配符展开导致意外行为。该建议未被标记为 resolved，但 PR 最终获得批准，表明团队可能接受当前简化写法或将在后续迭代中修复。
- Shell 脚本健壮性建议：启用 `set -u` 和引用变量 (correctness)：该建议未被作者采纳或回应，但 PR 最终获得批准，可能团队认为测试脚本在受控环境下可接受当前写法。

## 风险与影响

- 风险：
  - 外部依赖风险：工作流依赖 SUBMITTER\_URL（RayJob 提交器服务）和预构建的 Docker 镜像 IMAGE。如果这些基础设施不可用或版本不匹配，工作流将失败，且难以通过单元测试覆盖。
  - Shell 脚本健壮性：测试脚本未启用 `set -u`，未引用变量可能因通配符导致意外行为；未设置 `pipefail` 可能掩盖管道错误。这些在非预期环境中可能导致误通过或难以调试的失败。
  - 环境限制：工作流仅在特定 runner (linux-aarch64-a2b3-8) 上运行，该 runner 可能不具备通用性，在其他环境无法直接复现。
- 影响：
  - 用户影响：无。不修改任何运行时代码或 API。

- 系统影响：增加一个 CI 工作流，在 main 分支 push 和 PR 时触发（限定路径），增加 CI 资源消耗，但仅针对 Ascend 相关变更。
- 团队影响：为 Ascend CI 维护者提供了多节点测试覆盖，有助于提前发现多节点兼容性问题。
- 风险标记：外部依赖，Shell 脚本健壮性，环境限制

## 关联脉络

- 暂无明显关联 PR