

# PR #6275 完整报告

verl-project/verl

[ci] chore: slim down TRT-LLM CI

合并时间: 2026-05-08 19:34

原文链接: <http://prhub.com.cn/verl-project/verl/pull/6275>

## 执行摘要

- 一句话: 精简 TRT-LLM CI, 合并测试任务并移除冗余 job
- 推荐动作: 值得快速合并, 属于基础 CI 清理工作, 无功能性变更。建议后续关注 FSDP + TRT-LLM 的测试覆盖是否需单独补充, 以及观察拆分后的单元测试是否仍有稳定性问题。

## 功能与动机

per reviewer feedback, simplify the TRT-LLM CI: merge trtllm\_async\_unit\_tests into trtllm\_unit\_tests by adding test\_trtllm\_abort.py to the existing pytest invocation... Drop the e2e\_grpo\_trainer\_fsdp-qwen2 job; FSDP coverage with TRT-LLM rollout is no longer required (Megatron variants remain). Saves one runner per workflow run and consolidates trtllm unit tests into a single job with broader trigger coverage.

## 实现拆解

1. 合并单元测试 job: 在 e2e\_fully\_async\_policy\_trtllm.yml 中删除原有的 trtllm\_async\_unit\_tests job, 将 test\_trtllm\_abort.py 的测试步骤添加到 trtllm\_unit\_tests job 中 (在 e2e\_ppo\_grpo\_trainer\_trtllm.yml 的 pytest 调用后新增 `pytest -v -s --durations=20 tests/workers/rollout/rollout_trtllm/test_trtllm_abort.py`) 。
2. 移除 FSDP E2E job: 删除 e2e\_ppo\_grpo\_trainer\_trtllm.yml 中的 e2e\_grpo\_trainer\_fsdp-qwen2 job 及其所有步骤, 因为 FSDP 与 TRT-LLM 的测试覆盖不再需要, Megatron 变体已足够。
3. 清理触发路径: 从 e2e\_ppo\_grpo\_trainer\_trtllm.yml 的 pull\_request 路径列表中移除 examples/grpo\_trainer/run\_qwen3\_8b\_fsdp.sh, 避免该文件变更触发不必要的 CI 运行。
4. 更新依赖列表: 在 e2e\_fully\_async\_policy\_trtllm.yml 的 cleanup job 的 needs 中移除已删除的 trtllm\_async\_unit\_tests job。
5. 修复合并后的测试冲突: 在第二个 commit 中, 由于 test\_trtllm\_abort.py 与 test\_async\_server.py 等在同一个 Python 进程中执行时因反复 ray.init/ray.shutdown 加载 TRT-LLM 模型导致 SIGSEGV, 因此将测试拆分为两个独立的 pytest 调用。

关键文件:

- .github/workflows/e2e\_ppo\_grpo\_trainer\_trtllm.yml (模块 CI 配置; 类别 infra; 类型 infrastructure): 核心变更文件: 合并 trtllm\_unit\_tests 和 e2e\_grpo\_trainer\_fsdp-qwen2 job, 删除 54 行代码, 新增 3 行, 影响单元测试和 E2E

测试配置。

- `.github/workflows/e2e_fully_async_policy_trtllm.yml` (模块 CI 配置; 类别 `infra`; 类型 `infrastructure`): 辅助变更文件: 删除 `trtllm_async_unit_tests` job, 更新 `cleanup` job 的依赖关系, 删除 33 行代码, 新增 1 行。

关键符号: 未识别

## 评论区精华

无 review 评论。

- 暂无高价值评论线程

## 风险与影响

- 风险:
  1. 测试稳定性风险: 首次合并 `test_trtllm_abort.py` 到同一个 `pytest` 进程时出现 `SIGSEGV (exit 139)`, 已在第二次 commit 拆分为单独调用得到解决, 但需关注未来是否还有其他跨文件测试冲突。
  2. 覆盖完整性风险: 移除了 `e2e_grpo_trainer_fsdp-qwen2` job, 如果后续 `FSDP + TRT-LLM` 的兼容性出现回归, 可能无法及时发现。PR 说明 `Megatron` 变体已足够覆盖, 但长期仍需确保其他 `FSDP` 场景 (如其他模型) 的测试。
- 影响:
  1. CI 资源节省: 每次 workflow 运行减少一个 runner, 预计减少约 30-40% 的 `TRT-LLM` CI 计算成本。
  2. 测试覆盖维持: 单元测试仍覆盖了 `trtllm_abort`、`async_server`、`inter_node_rollout`、`rollout_utils` 等关键模块; E2E 测试保留了 `Megatron` 变体, 基本覆盖了主要用例。
  3. 维护简化: 合并后只需维护一个单元测试 job, 触发路径更清晰。 - 风险标记: 缺少测试覆盖, `SIGSEGV` 历史问题

## 关联脉络

- PR #5631 [rollout] feat: enable Async RL for trtllm rollout: 与本次修改的 `trtllm` 异步测试相关, 引入了 `test_trtllm_abort.py` 等异步测试文件。
- PR #6230 [rollout] fix: trtllm rollout docker image and a few scripts: 同样涉及 `TRT-LLM` CI 配置调整, 与本次精简工作属于同一维护线。