

# PR #1729 完整报告

THUDM/slime

[docker] update megatron

合并时间: 2026-03-18 10:32

原文链接: <http://prhub.com.cn/THUDM/slime/pull/1729>

## 执行摘要

本 PR 通过更新 Docker 中的 SGLang 补丁和修改路由器配置，修复了 RDMA 传输超时导致的会话错误中断问题。变更影响 CI 镜像和部署稳定性，建议相关团队关注配置调整，以提升系统在高负载下的可靠性。

## 功能与动机

动机源自提交消息 "remove failed\_sessions.add to prevent rdma session from breaking"，旨在处理高负载下 PCIe 争用引起的 RDMA 超时，避免误判服务器死亡而中断会话。这解决了内部 bug，确保会话处理更稳健。

## 实现拆解

- Docker 模块: 更新 `docker/patch/latest/sglang.patch`，删除 `failed_sessions.add` 和相关错误日志，优化 MooncakeKVManager 中的失败处理。同时升级 `docker/version.txt` 至 `nightly-dev-20260318a`。
- Rollout 模块: 修改 `slime/ray/rollout.py` 中的 `_start_router` 函数，添加对 PD disaggregation 的参数支持并设置 `disable_circuit_breaker` 为 `True`，注释解释: "Disable circuit breaker to prevent RDMA transfer timeouts from marking decode workers as dead. Timeouts are transient (PCIe contention under high load) and do not indicate a dead server."
- 清理模块: 移除 `slime/rollout/generate_hub/` 下的未使用文件，简化代码库。

## 评论区精华

无 review 评论，变更直接合并，表明团队信任作者或变更较紧急。

## 风险与影响

- 风险: 补丁更新可能引入上游 SGLang 新 bug；禁用断路器可能掩盖真正服务器故障；会话失败逻辑调整影响错误恢复。具体风险点在 `sglang.patch` 删除代码和 `rollout.py` 配置变更。
- 影响: Docker 镜像需更新以获取修复，影响 CI 流程；路由器行为变更改善高负载稳定性，但需监控潜在问题；移除文件对功能无影响。

## 关联脉络

关联 PR 1743 (类似 Docker 更新)、1751 和 1765 (修改同一 `rollout.py` 文件), 显示近期趋势是持续优化路由器配置和修复 RDMA 相关 bug, 反映团队在提升分布式系统稳定性的努力。