PR #21299 完整报告

sgl-project/sglang

[Disagg DP-Balancing] Refactor Disagg Conn and Fix Hang with total_request/total_tokens Balancing

合并时间：2026-03-31 18:01

原文链接：http://prhub.com.cn/sgl-project/sglang/pull/21299

执行摘要

此 PR 通过重构 disaggregation 接收器生命周期，修复了因 DP 平衡（如 total_requests/total_tokens）导致的 hang 问题，确保 decode ranks 间状态同步，提升系统可靠性。

功能与动机

此变更旨在解决 issue #21297 报告的 PD disaggregation hang 问题。当 prefill 使用内部 DP 平衡时，decode 可能在 bootstrap_room -> prefill_dp_rank 映射可用前收到请求，导致各 rank 独立解析 pending 请求，引发不一致的本地队列状态和 handshake stall。PR body 中引用 issue #21680 进行讨论，最终采用 receiver 生命周期重构方案。

实现拆解

关键改动按模块梳理：

- 基础层（base/conn.py）：添加抽象方法 init() 和 send_metadata()，分离初始化和元数据发送。
- 通用层（common/conn.py）：重构 CommonKVReceiver，__init__ 仅设置基本属性（如 bootstrap_addr），init() 方法解析 prefill DP-rank 并调用 _setup_bootstrap_infos()，延迟 bootstrap 设置。
- 解码层（decode.py）：修改 add() 方法，立即创建 receiver 并加入队列，但延迟调用 init()；pending 请求列表改为存储 DecodeRequest 对象；移除 leader-based resolution 逻辑。
- 具体实现层（fake/mooncake/mori/nixl/conn.py）：适配新方法，移除 prefill_dp_rank 参数，添加 init() 和 send_metadata() 实现。
- 测试层（test_disaggregation_dp_attention.py）：新增 TestDisaggregationDPAttentionTotalRequests 和 TestDisaggregationDPAttentionTotalTokens 测试类，覆盖 DP 平衡路径。

评论区精华

review 讨论聚焦以下要点：

- 正确性问题：ShangmingCai 指出“If we only let the leader rank do the bootstrapping and broadcast to other ranks, will other ranks fail to init prefill_info_table?”，

weireweire 回应修改为各 rank 自行运行但添加 sync 逻辑。

- 设计权衡：ShangmingCai 建议“Maybe we should only call decode_req.kv_receiver.abort() here... to avoid double free”，weireweire 表示 failure handling 未在此 PR 解决，需后续决策。
- 演进方向：基于 issue #21680 讨论，从 leader-based workaround 转向 receiver 生命周期重构，确保一致性。

风险与影响

风险：

1. 新生命周期分离增加复杂性，若 init() 或 send_metadata() 实现不一致，可能导致 handshake 失败。
2. failure 处理逻辑不完整（如 abort 问题），可能引发资源泄漏或错误报告不及时。
3. 同步机制依赖 ranks 间状态一致，若广播失败，可能重现 hang 问题。
4. 测试覆盖虽增强，但可能遗漏故障场景的 edge cases。

影响：

- 对用户：修复 hang 问题，提升使用 DP 平衡时的服务稳定性。
- 对系统：改善 decode 侧 pending 请求处理，防止因状态不一致导致的 stall。
- 对团队：需适应新 receiver 生命周期，failure handling 需后续工作。

关联脉络

与历史 PR 关联较弱，主要基于 issue #21680 的讨论进行重构。近期历史 PR 中涉及 disaggregation 的测试优化（如 PR 21745），但未直接关联核心逻辑变更。此 PR 揭示了 disaggregation 模块中 DP 平衡路径的演进，强调一致性同步的重要性。