

PR #2484 完整报告

radixark/miles

docs: add disaggregated RL rollout guide

合并时间: 2026-08-13 06:39

原文链接: <http://prhub.com.cn/radixark/miles/pull/2484>

执行摘要

- 一句话: 新增分离式 RL rollout 架构指南, 同步 6 个文档页面
- 推荐动作: 值得精读, 尤其 Topologies 表、--rollout-external 的定位说明、disk-delta 5 步生命周期与 policy-version 契约部分。对于要接入或实现外部 rollout 服务的工程师, 这份文档就是接口契约草案; 对纯功能使用者, 重点看 Quick Start 与 Training Backends 中被同步修改的 --update-weight-transfer-mode 说明。设计上值得借鉴的是文档先定义契约、再等实现落地的先行方式, 以及 commit 序列中对外部项目表述的克制处理。

功能与动机

训练与 rollout 的资源画像完全不同: 训练 ranks 构成长期存续、紧耦合的集群, 而 rollout 需求是突发性的, 可以在全球范围内动态扩容 GPU。PR body 明确说明: Making rollout an independent service lets miles add or remove inference capacity without resizing the trainer, place rollout where GPUs are available, isolate replica failures, and keep fully async generation supplied。同时这种分离要求 RL 特有的正确性契约——miles 必须发布不可变的 policy 版本、在 rollout 请求上表达版本需求、并记录每条返回轨迹由哪个版本生成。指南还澄清了 --rollout-external 的定位 (挂接外部引擎而非外部服务), 并把 disk-delta 定位为当前的发布基础。

实现拆解

1. 新增核心指南页: docs/advanced/disaggregated-rollout.md (+332 行)。以训练与 rollout 资源画像差异开篇论证解耦价值; 用三行拓扑表区分 miles-managed rollout / Attached SGLang engines / External rollout service, 明确前两种当前已支持、第三种 coming soon; 给出独立 GPU 布局示例 (--actor-num-nodes 1、--rollout-num-gpus 8、--rollout-num-gpus-per-engine 2) 与外部引擎挂接参数 (--rollout-external、--rollout-external-engine-addr), 并强调 --rollout-external 不会移交权重更新所有权。
2. 梳理权重同步路径: 用表格对比 --update-weight-transfer-mode 三档——broadcast (NCCL 默认)、p2p (RDMA 直写, 集群内优化)、disk-delta (共享存储发布版本化增量); 随后给出 disk-delta 的 5 步发布 / 激活生命周期 (CPU snapshot → 字节对比 → 原子发布 weight_vNNNNNNN / → 本地物化与校验 → 暂停 reload 激活), 并说明 XOR 与 overwrite 两种编码的取舍 (XOR 紧凑但只能应用一次, overwrite 更大但幂等)。
3. 注册导航并联动既有页面: docs/docs.json 在 Advanced Features > Scale & Reliability 分组插入新页面 (置于 p2p-weight-transfer 与 disk-offload 之间);

docs/advanced/index.md 新增指南卡片并更新段首特性列表；pd-disaggregation.md 增加范围区分互链；quick-start.md、training-backend.md、fully-async.md 三处将权重传输描述从二分（broadcast/p2p）扩展为三分（含 disk-delta），并把广播表述改为中性的按配置模式同步。

4. 迭代打磨与验证：18 个提交完成多轮内容收敛（详见评论区精华）；验证手段为 git diff --check、python -m json.tool docs/docs.json 与全量内部链接检查，未引入自动化测试。

关键文件：

- docs/advanced/disaggregated-rollout.md（模块 高级指南；类别 docs；类型 documentation）：新增 332 行的核心指南页，定义了三种 rollout 拓扑、三种权重同步模式、disk-delta 发布激活生命周期与 policy-version 契约，是本次变更的主体与后续实现的契约草案。
- docs/docs.json（模块 站点导航；类别 config；类型 configuration）：文档站点导航配置，将新页面注册到 Advanced Features > Scale & Reliability 分组，是站点结构生效的关键配套改动。
- docs/advanced/index.md（模块 高级入口；类别 docs；类型 documentation）：Advanced 功能入口页，新增 Disaggregated RL Rollout 卡片并更新段首特性列表，保证高级文档入口与新增指南一致。
- docs/user-guide/fully-async.md（模块 用户指南；类别 docs；类型 documentation）：Fully Async RL 流程文档需要同步权重更新描述，将广播改为按配置的更新模式同步，并保持 pause-generation-mode 三态说明。
- docs/getting-started/quick-start.md（模块 快速上手；类别 docs；类型 documentation）：快速上手是高流量入口，把 --update-weight-transfer-mode 从二分（broadcast/p2p）扩展为三分（含 disk-delta）并补充链接，避免初学用户漏掉第三种模式。
- docs/user-guide/training-backend.md（模块 训练后端；类别 docs；类型 documentation）：Training Backends 文档与本 PR 的 GPU 布局讨论直接相关，同步更新权重传输模式说明并链接 disk-delta 到新指南。
- docs/advanced/pd-disaggregation.md（模块 阶段拆分；类别 docs；类型 documentation）：PD 拆分指南与新的 Disaggregated RL Rollout 指南在概念上容易混淆，增加互链与范围区分说明，明确前者是部署内部阶段分离。

关键符号：未识别

评论区精华

该 PR 没有任何 inline review 评论（comments_count = 0、review_comments_count = 0），唯一正式审核是 Zhichenzzz 的空正文 APPROVED。可观察的讨论体现在 18 个提交的演进序列中：

1) 对第三方参考实现 Stitch (Modal 团队) 的定位经历 identify Stitch reference integration → avoid priority claim for Stitch → keep Stitch mention focused → link Stitch's miles-fork notes from the rollout-service contract 四轮收敛，说明作者刻意控制表述分寸，避免读者把外部项目误读为官方路线图优先级。2) 测量数据经历 add Kimi K2.6 weight-sync measurement → simplify weight-update measurements → focus rollout

weight-sync measurements → tighten rollout measurement context 多轮瘦身，最终只保留与 rollout 权重同步直接相关的上下文。3) use lowercase miles in rollout guide 统一了产品名拼写规范；合并阶段由 Zhichenzzz 补入对 Nan Jiang / Modal 团队与 Jason Mancuso 的致谢，并把 Stitch 的 miles-fork 笔记互链进外部服务契约段落。

- 暂无高价值评论线程

风险与影响

- 风险：1) 文档与 CLI 契约漂移：指南密集出现 `--rollout-external`、`--rollout-external-engine-addr`s、`--update-weight-disk-dir`、`--update-weight-local-checkpoint-dir` 及 XOR/overwrite 等开关与编码名，若对应实现尚未合入或命名调整，本指南会成为误导源；PR 把外部服务标注为 coming soon 已降低此风险，但实现侧需跟进对齐。2) 数据时效性：Kimi K2.6 权重同步测量涉及具体数字，模型规模与网络环境变化后容易过期。3) 无自动化验证：仅依赖 json.tool 与链接检查，docs/docs.json 结构与页内锚点（如 training-backend#3-choosing-the-gpu-layout）后续改动可能失配。4) 影响面：7 个文件、6 个页面联动改文案，其中 Quick Start 与 Training Backends 是高频入口，措辞变化需与真实行为保持一致。
- 影响：对读者（用户 / 工程师）：首次获得跨集群、跨 region 解耦 rollout 的完整指南，明确了三种拓扑与三种权重同步模式的选择依据，以及 `--rollout-external` 的真实语义。对团队：以文档先行方式锁定 policy-version 契约与包中立边界的术语体系，为后续外部 rollout 服务实现提供设计锚点与接口草案。对文档站点：新增一个 Advanced 页面并联动 6 页，Advanced Features > Scale & Reliability 分组导航路径发生变化。整体影响范围中等偏上（对 rollout 功能线是奠基性文档），但无运行时影响，风险集中在文档与未来实现的一致性上。
- 风险标记：文档与 CLI 契约漂移风险，coming-soon 预告内容，测量数据易过期，无自动化测试覆盖

关联脉络

- PR #2375 docs: reorganize the Training Backends section, drop the experimental FSDP framing: 两 PR 均改动 docs/user-guide/training-backend.md 与 docs/docs.json；本指南引用了其 Choosing the GPU layout 锚点，属于同一文档结构演进线。
- PR #2278 feat(session): request and assemble additional R3 rows under in-place weight updates: 描述 in-place 权重更新下多轮 rollout 的增量请求与拼接，与本指南中的 pause/update/resume 生命周期及权重同步语义直接相关。
- PR #2369 fix(rollout): normalize rewards per rollout: rollout 侧奖励归一化修复，与本指南强调的记录每条轨迹由哪个 policy 版本生成的 rollout 数据正确性主线互相印证。
- PR #2478 docs: rewrite INT4 QAT guide: 同为 docs/advanced 下的指南级文档维护，反映该区域持续的文档沉淀模式。