

PR #2376 完整报告

radixark/miles

docs(developer): rewrite the developer guide against the code

合并时间: 2026-08-12 05:50

原文链接: <http://prhub.com.cn/radixark/miles/pull/2376>

执行摘要

- 一句话: 按代码重写开发者文档并新增版本页
- 推荐动作: 值得精读, 尤其适合两类读者: 新贡献者应细读 contributor-guide 的 Running CI 与 code style 章节, 按文档步骤验证测试注册; 维护者应阅读 versions.md 的 bumping principle 与镜像标签规则, 这是全仓库版本管理策略第一次被系统化写成文字。可以借鉴的做法: 文档中每个结论都标注来源(代码文件、workflow、.claude), 并将“如何验证”写进文档(如 run_suite.py --list-only、dev 标签时间戳), 值得在后续文档维护中推广。

功能与动机

PR body 明确指出旧文档“actively mislead a new contributor”: 它记录了项目并未使用的 black 行宽 100、已随 #2272 消失的 backends/experimental/fsdp_utils/ 目录, 以及 arguments.py 中从未存在的 `--debug-determinism`、`--debug-weight-sync`、`--debug-rollout-print-every` 等 flag; 同时缺少贡献者最需要的“如何让 CI 运行我的测试”和“依赖版本如何配合”两件事。重写目标就是让每一句陈述都能从代码、workflows 或 .claude 中找到依据。

实现拆解

1. 重写贡献指南 contributor-guide.md: 从仓库树重新生成 Repository layout (移除 backends/experimental/, 补充 dashboard、true_on_policy、miles_plugins/optimizers/); 代码风格改为基于 .pre-commit-config.yaml 的 hook 说明, 纠正 black 行宽为 119, 并首次记录三条 pygrep 禁令 (mpu.get_、裸 AutoConfig/AutoTokenizer、huggingface-cli) 及替代 API; 新增 .claude 目录说明与 doc-dev sentinel 表; 大量补充 Running CI 章节, 包括 tests/fast/ 自动注册、register__ci 的 AST 解析机制、run_suite.py --list-only 验证、各 run-ci-* label 含义, 以及首次文档化的 PR-description CI tags (ci-image-tag、ci-sglang-pr、ci-megatron-pr)。
2. 重写调试文档 debug.md: 删除不存在的 debug flags 和信号短语, 改为从代码中提取的真实调试面: `--debug-rollout-only` / `--debug-train-only` (互斥断言)、`--save/load-debug-rollout-data`、`--ci-inject-rollout-data-path` (保留引擎并在丢弃前比对生成结果, min-match-ratio 默认 0.9)、`--debug-skip-weight-update`、`--debug-disable-optimizer`、`--debug-exit-after-rollout`、`--debug-deterministic-collective` (det_nccl 固定树序归约); 并逐一文档化 `--ci-test` 断言 harness 的每个 checker 及其阈

值、逃生舱口和 `--ci-save/load-grad-norm`、`--ci-save/check-model-hash` 等参数。

3. 新增 `versions.md`: 说明 `miles`、`sglang-miles` 分支、`miles-main` 分支三棵可编辑安装源码树; 列出 `docker/Dockerfile build-args` 默认值 (`SGLANG_IMAGE_TAG v0.5.16`、`SGLANG_COMMIT` 空等)、`build.py` 变体表、`requirements.txt` 内联 `pin` 约定、TE 2.17.0 三元组断言与 `cu13` 补丁; 整理发布镜像与标签规则、`dev` 标签的 24h `check-upstream` 自动构建机制; 总结 `bumping principle`: 只改 `pin` 所在处一次、优先移动分支、镜像相关 `bump` 必须在对应 PR 内验证 (`pr-` 镜像高于 `ci-image-tag`)、`ref bump` 用 PR 描述指令验证, 并给出 `cuDNN` 降级 (`CUDNN_STATUS_BAD_PARAM`)、TE 补丁失效等失败模式表。
4. 修正 `architecture.md` 与 `index.md`, 删除 `migration.md`, 更新 `docs.json` 导航: 包树重新生成 (`backends/experimental/` 移除、`fsdp_utils/` 与 `megatron_utils/` 并列、`miles_plugins/` 四子目录、测试树新增 `manual/` 与 `snapshots/`); “Where common changes go”增加 `FSDP` 行, `weight sync` 指向 `megatron_utils/update_weight/` 或 `fsdp_utils/update_weight_utils.py`; 删除记录 `sync`→`async` 旧迁移的 `migration.md` (同步从导航和卡片移除); `docs.json` 将 `migration` 页替换为 `versions` 页。
5. 配套调整: 仅涉及文档导航配置 (`docs.json 2+/2-`) 与索引卡片文案, 无源码、测试或部署改动; PR body 明确说明未给 `/developer/migration` 加 `redirect`, 留待 #2375 的 `redirects` 块落地后补。

关键文件:

- `docs/developer/contributor-guide.md` (模块 贡献指南; 类别 docs; 类型 documentation): PR 核心文件: 重写仓库布局、代码风格 `hook` 说明、`.claude` 解释和完整的 CI 运行指引 (注册机制、`label`、PR-description tags), 是修复误导性内容的重点。
- `docs/developer/debug.md` (模块 调试指南; 类别 docs; 类型 documentation): 纠错重点: 删除三个不存在的 `debug flags` 和虚假日志信号, 替换为真实调试面 (`debug-*only`、`debug data` 读写、`ci-inject-rollout-data-path`、`det_nccl`) 以及 `--ci-test` 断言 `harness` 各 `checker` 的阈值与逃生舱口。
- `docs/developer/versions.md` (模块 版本说明; 类别 docs; 类型 documentation): 全新增页面: 首次系统说明三棵源码树 (`miles/sglang-miles/miles-main`) 的关系、全部版本 `pin` 的存放位置、镜像标签规则以及 `bumping principle`, 是依赖升级与镜像维护的权威参考。
- `docs/developer/migration.md` (模块 迁移指南; 类别 docs; 类型 deletion): 整页删除: `sync`→`async` 迁移早已完成、`breaking flags` 列表仅剩一条记录, 保留会继续误导读者; 同时从导航和索引卡片移除。
- `docs/developer/architecture.md` (模块 架构说明; 类别 docs; 类型 documentation): 修复包树中自 #2272 起不再存在的 `backends/experimental/fsdp_utils/` 目录描述, 补充 `miles_plugins/` 树、`FSDP weight sync` 路径、测试树 `manual/` 与 `snapshots/`, 并修正 `weight sync` 的指引目录。
- `docs/docs.json` (模块 文档导航; 类别 config; 类型 configuration): 站点导航配置: `Developer Guide` 页面列表移除 `migration`、加入 `versions`, 保证文档站与 PR 内容一致。
- `docs/developer/index.md` (模块 文档索引; 类别 docs; 类型 documentation): 索引卡片随内容重写调整: 移除 `Migration Guide` 卡片、新增 `Versions and Images` 卡片, 并同步

卡片文案与新手步骤。

关键符号：未识别

评论区精华

Review 由 Shi-Dong 完成，共 5 条评论后 APPROVED (LGTM!)，全部在最后提交中处理：

- architecture.md：对“reached by name from a flag”的表述提出疑问“I'm not sure what 'reached by name from a flag' means.”，最终改为更直白的“a plugin is loaded only when a run names its import path in a flag”。
- debug.md：认为 page description“过于详细”，建议“Probably just replace it with 'Useful tips for debugging'”，已采纳。
- debug.md：在 --debug-rollout-only 行建议“Add FSDP alongside Megatron?”，已改为“The training backend (Megatron or FSDP) is never initialized”。
- versions.md：两处大小写规范“miles -> Miles”，均已修改（标题与 three-trees 表格）。
- 无未解决疑虑；PR body 中作者主动指出与 #2375 的 docs.json / developer 页冲突需合并时解决，以及 contributor-guide 与 docs/ci/contributor-guide.md 的 CI 内容重叠留作后续。
- miles_plugins 加载方式表述不清 (documentation)：最终提交改为直白表述：“a plugin is loaded only when a run names its import path in a flag (--spec, or one of the --custom-*-path flags)”。
- debug 页 description 过于详细 (style)：已采纳，page description 缩短为“Useful tips for debugging.”。
- debug-rollout-only 描述补充 FSDP (documentation)：已修改为“The training backend (Megatron or FSDP) is never initialized.”。
- versions.md 标题大小写 (miles -> Miles) (style)：已修改为“Miles, SGLang, and Megatron-LM”。
- versions.md 正文大小写 (miles -> Miles) (style)：已修改为“Miles”。

风险与影响

- 风险：纯文档变更，无代码回归风险，但存在以下具体风险：
 - 版本信息漂移：versions.md 记录了 Dockerfile build-args 当前默认值（如 SGLANG_IMAGE_TAG v0.5.16、WHEELS_TAG cu130-x86_64 等）和 dev 标签更新节奏，这些内容会随构建配置变化而过时，需要后续维护者同步更新。
 - 删除页面无重定向：migration.md 被直接删除且未设置 redirect，若外部或历史链接指向 /developer/migration 将 404；PR body 说明等 #2375 落地后补，存在窗口期。
 - 与 #2375 合并冲突：两个 PR 都改动 docs.json 和 docs/developer/index.md，且 #2375 还触及 architecture.md，后合入方需要解决冲突，否则可能破坏导航 JSON。
 - CI 约定变更：contributor-guide 首次固化了 run-ci-* label、PR-description tags、注册机制等行为描述，若 CI 工作流后续调整而文档未同步，会再次造成误导。

- 影响：影响范围集中在开发者体验与文档站结构：新贡献者 onboarding 路径（仓库布局、风格、.claude、CI 触发）从错误信息修正为可执行指南；调试文档从虚假 flag 变为真实可用的调试面，有助于快速定位 rollout/training 问题；新增 versions.md 让涉及依赖升级、镜像构建的维护者有单一依据来源。对团队而言，文档导航减少了一个页面（migration）、新增一个页面（versions），Developer Guide 卡片顺序调整；对 CI 使用有直接帮助（PR 标签、注册验证）。影响程度中等偏上，因为它是新贡献者的第一手资料，但无运行时影响。
- 风险标记：文档与代码漂移风险，删除页面无重定向，与 #2375 合并冲突

关联脉络

- PR #2386 fix: drop context parallelism from the FSDP backend: FSDP 后端结构调整，architecture.md 中 fsdp_utils/ 的描述需要与 #2386 之后的代码保持一致（同属 FSDP 目录与能力描述演进）。
- PR #2384 fix(fsdp): stop store_true from shadowing bool defaults in FSDPArgs: FSDP 参数行为修复，与 debug.md 中 FSDP 相关 flag 的准确描述相关联。
- PR #2279 Run the launch script snapshot tests by hand instead of in CI: 将测试移入 tests/manual/ 并保留 tests/snapshots/，architecture.md 测试树中新增 manual/ 与 snapshots/ 的描述即源于此。
- PR #2374 docs: split FAQ out of Resources and link the blog to LMSYS: 同样调整 docs.json 导航结构，与 #2376 并行的文档站信息架构演进。
- PR #2303 feat(ci): add authorized Neon SQL workflow: CI workflow 新增与文档化，与 contributor-guide 中 Running CI 章节同属 CI 使用说明线。