

# PR #6353 完整报告

verl-project/verl

[doc] chore: verl Ascend doc refactor

合并时间: 2026-05-14 22:02

原文链接: <http://prhub.com.cn/verl-project/verl/pull/6353>

## 执行摘要

- 一句话: Ascend 文档目录重构与内容优化
- 推荐动作: 建议阅读本 PR 的文档结构变化以了解当前 Ascend 文档的组织方式, 但不建议直接引用其中的路径和配置示例 (因存在已知错误)。后续应追踪修复评论中提出的问题, 并为删除的文档设置重定向或通知。

## 功能与动机

PR body 仅说明 'verl Ascend doc refactor', 但从变更内容看, 主要目的是将过去散落在 examples、profiling 等目录的文档统一归入 model\_support、dev\_guide 等结构化目录, 新增 README 以提供清晰的导航, 并淘汰过时内容 (如 run\_qwen3\_32B\_megatron\_1k\_256k\_npu.md)。此前有多项 NPU 文档 PR (如 #6337、#6347) 已逐步调整, 本次重构是这些工作的集中整合。

## 实现拆解

1. 新增顶层导航文件: 在 docs/ascend\_tutorial/ 下新增 README.md, 列出快速开始、特性支持、模型支持、开发指南等模块的目录结构及其链接。
2. 目录重组: 将原位于 examples/ 下的多个最佳实践文档 (ascend\_sglang\_best\_practices.rst、ascend\_retool\_best\_practice.rst、dapo\_multi\_model\_optimization\_practice.md、gspo\_optimization\_practice.md) 移至 model\_support/examples/; 将 profiling/ 下的两个 profiling 文档 (ascend\_profiling\_zh.rst、ascend\_profiling\_en.rst) 和 ascend\_performance\_analysis\_guide.md 移至 dev\_guide/performance/; 删除 examples/run\_qwen3\_32B\_megatron\_1k\_256k\_npu.md 和 feature\_support/ascend\_consistency.rst。
3. 新增评测指南: 创建 dev\_guide/model\_dev/evaluation.md, 详细说明使用 AISBenchmark 进行模型评测的步骤 (包括 vLLM 和 sglang 后端)。
4. 更新全局索引: 修改 docs/index.rst, 移除已删除文件的条目, 调整路径以匹配新目录结构。
5. 内容微调: 在 ascend\_sglang\_best\_practices.rst 中修正了脚本链接的 commit id、更新了硬件表格、优化了部分描述。
6. 未修正的已知错误: review 指出的反斜杠路径、端口不匹配、设备索引超限、URL 指向错误等问题在合并版本中仍存在。

关键文件：

- docs/ascend\_tutorial/dev\_guide/model\_dev/evaluation.md（模块 开发指南；类别 docs；类型 documentation）：新增模型评测指南，是本次重构的核心新增内容，但 review 发现两处 high priority 错误。
- docs/ascend\_tutorial/README.md（模块 入口文档；类别 docs；类型 documentation）：新增的顶层 README，作为整个 Ascend 教程的目录入口，对用户导航至关重要。
- docs/ascend\_tutorial/model\_support/examples/ascend\_sglang\_best\_practices.rst（模块 最佳实践；类别 docs；类型 rename-or-move）：被移动并修改内容（更新引用脚本的 commit id 等），但 review 也发现了设备索引和 URL 问题。
- docs/index.rst（模块 全局索引；类别 docs；类型 documentation）：必须更新以反映所有文件的新路径和删除，是文档可被 sphinx 正确构建的关键。

关键符号：未识别

## 评论区精华

Gemini Code Assist 在 review 中发现了 4 个高优先级问题，均未得到作者回应即被合并：

- evaluation.md 第 33 行的链接使用反斜杠和错误相对路径，应改为 `./../advance/checkpoint.rst`。
- evaluation.md 第 84 行的 client 配置端口 (8080) 与服务端端口 (6380) 不一致。
- ascend\_sglang\_best\_practices.rst 第 135 行的 ASCEND\_RT\_VISIBLE\_DEVICES 包含 9 个索引 (0-8)，但 NPUS\_PER\_NODE=8。
- ascend\_sglang\_best\_practices.rst 第 291 行的 profiling 链接指向 **volcengine** 组织且路径过时。
- evaluation.md 中的路径和端口错误 (correctness)：建议使用正斜杠相对路径，并统一端口为 6380。
- ascend\_sglang\_best\_practices.rst 中的设备索引和 URL 错误 (correctness)：建议修正设备索引范围，并更新为正确的相对链接。

## 风险与影响

- 风险：文档变更本身风险较低，但 review 指出的 4 个错误均为 high priority 且未被修正，直接降低了文档的可靠性和可用性：
  - 反斜杠路径在非 Windows 环境下无效，可能导致用户无法跳转到参考文档。
  - 端口不匹配会导致评估流程失败，用户需要自行调试发现不一致。
  - 设备索引溢出可能导致训练 / 评测脚本因索引越界报错。
  - 错误 profiling 链接（指向 volcengine 而非 verl-project）使用户无法到达预期页面。
  - 删除的文档（run\_qwen3\_32B\_megatron\_1k\_256k\_npu.md）如果被外部引用，将出现 404。
- 影响：影响范围：所有通过网站或 GitHub 查阅 verl Ascend 文档的用户。目录重构后用户需要适应新路径，但 README 的加入提供了良好指引。训练示例和评测指南的整合使

文档更聚焦，但存在的内容错误可能使用户信任度下降。对系统无影响。

- 风险标记：文档路径错误易导致 404，端口不一致使评测流程失败，设备索引超限，URL 指向错误仓库

## 关联脉络

- PR #6337 [doc] chore: split install guidance and quickstart: 同属 Ascend 文档重构系列，此前已拆分安装与快速入门指南，本 PR 进一步将其他文档归入统一目录结构。
- PR #6347 [doc] refactor: added the document that collects statistics on models and algorithms that support the NPU: 新增模型支持文档，本 PR 将其链接纳入 README 并调整了 examples 路径。
- PR #6339 [doc] chore: add npu advanced features: 新增 NPU 高级特性文档，本 PR 也调整了其目录位置。
- PR #6328 [doc] refactor: Ascend docs rectification, add new FAQ questions: 此前已对 FAQ 进行修正，本 PR 是更大范围的结构性重构。