

# PR #23329 完整报告

sgl-project/sclang

Add IntraNode NVLink configuration in PD disaggregation docs

合并时间: 2026-05-14 14:23

原文链接: <http://prhub.com.cn/sgl-project/sclang/pull/23329>

## 执行摘要

- 一句话: 新增 Intra-Node NVLink 配置文档
- 推荐动作: 简单的文档更新, 值得阅读以了解新配置选项。

## 功能与动机

为了让用户了解如何利用 Intra-Node NVLink 进行 KV 缓存传输 (推荐用于 A100、H20、H100 等), 需在文档中补充对应的环境变量配置说明。PR body 提到用户可以通过设置 `SGLANG_MOONCAKE_CUSTOM_MEM_POOL=INTRA_NODE_NVLINK` 和 `MC_INTRANODE_NVLINK=true` 启动服务器。

## 实现拆解

1. 修改 docs/advanced\_features/pd\_disaggregation.md: 在第 133 行之后新增一段配置说明, 包含环境变量设置和注意事项。
2. 修改 docs\_new/docs/advanced\_features/pd\_disaggregation.mdx: 在相同位置添加相同内容, 保持新旧文档同步。
3. 修复拼写错误: 应 reviewer 要求将 `awared` 改为 `awarded`。
4. 解决冲突: 合并 main 分支后解决冲突。

关键文件:

- docs\_new/docs/advanced\_features/pd\_disaggregation.mdx (模块文档; 类别 other; 类型 core-logic): 新增 Intra-Node NVLink 配置说明, 与文档站点同步更新。
- docs/advanced\_features/pd\_disaggregation.md (模块文档; 类别 docs; 类型 documentation): 对应 Markdown 版本文档, 同样新增 Intra-Node NVLink 配置说明。

关键符号: 未识别

## 评论区精华

reviewer stmatengss 指出拼写错误 `awared` 应为 `awarded`, 作者已修复。其他讨论包括确认该功能在 Docker 镜像中不可用, 需从源码编译 Mooncake 并启用 IntraNode NVLink `cmake` 标志。

- 拼写错误修复 (style): 作者已修复。

- Docker 镜像可用性 (question): 作者确认不可用, 需从源码编译 Mooncake。

## 风险与影响

- 风险: 该 PR 仅为文档变更, 无技术风险。
- 影响: 影响用户: 提供正确配置方法, 提升使用体验。影响系统: 无。影响范围小。
- 风险标记: 暂无

## 关联脉络

- 暂无明显关联 PR