

PR #28078 完整报告

sgl-project/sglang

docs: add llm-d page under Advanced Features

合并时间: 2026-06-13 09:40

原文链接: <http://prhub.com.cn/sgl-project/sglang/pull/28078>

执行摘要

- 一句话: 新增 llm-d 集成文档页面
- 推荐动作: 该 PR 为常规文档更新, 内容清晰、变更简单, 无需深度精读。但可作为了解 SGLang 生态系统集成的参考。

功能与动机

SGLang 是 llm-d 支持的推理引擎之一, 该文档页帮助在集群规模下运行 SGLang 的用户从文档中发现 llm-d 集成能力。PR body 说明: 'llm-d coordinates a fleet of SGLang instances across a cluster so that performance holds up under real production traffic'。

实现拆解

1. 新建文档页面 docs_new/docs/advanced_features/llm-d.mdx: 包含标准 frontmatter (title、metatags.description), 正文介绍 llm-d 是什么、它为 SGLang 部署添加的集群级能力 (前缀感知路由、分布式 KV 缓存管理、预填充 / 解码分离、SLO 感知自动扩缩容、多引擎统一控制面) 以及快速开始指引。
2. 注册导航在 docs_new/docs.json 的 Advanced Features 分组中, 在 sgl_model_gateway 之后、deterministic_inference 之前插入 "docs/advanced_features/llm-d" 页面引用。
3. 验证作者提到已本地使用 mint dev 构建预览, mint broken-links 报告无断裂链接。

关键文件:

- docs_new/docs/advanced_features/llm-d.mdx (模块 文档; 类别 docs; 类型 content-addition): 新增的文档核心文件, 详细介绍 llm-d 集成能力, 是 PR 的主要变更。
- docs_new/docs.json (模块 文档配置; 类别 config; 类型 configuration): 导航配置文件, 新增 llm-d 页面注册, 确保页面在文档站点中可见。

关键符号: 未识别

评论区精华

无 review 评论。仅有一条 gemini-code-assist 自动消息提示日配额已用尽, 与 PR 内容无关。

- 暂无高价值评论线程

风险与影响

- 风险：纯文档变更，无技术风险。唯一潜在风险是新页面路径在 docs.json 中的位置可能导致已有书签失效，但仅新增条目，不影响其他页面路径。
- 影响：对用户：为需要集群规模部署 SGLang 的用户提供明确的 llm-d 集成指引和外部文档链接。对系统：无任何代码、性能或模型行为影响。对团队：增加了文档维护负担，但内容指向外部项目，更新频率低。
- 风险标记：纯文档变更

关联脉络

- 暂无明显关联 PR