

PR #22923 完整报告

sgl-project/sglang

docs: fix incorrect default max-payload-size in gateway config reference

合并时间: 2026-04-16 13:25

原文链接: <http://prhub.com.cn/sgl-project/sglang/pull/22923>

执行摘要

本次 PR 修正了 sgl-model-gateway 配置参考文档中 `--max-payload-size` 参数的默认值错误，将文档记载的 256 MB 改为代码实际默认的 512 MB，解决了因文档与代码不一致导致的用户混淆问题。变更仅涉及单行文档修正，无任何代码或测试改动，风险极低。

功能与动机

为什么需要这个变更？根据 PR body 描述，文档中 `--max-payload-size` 的默认值（256 MB）与代码实现（Python 的 `router_args.py` 和 Rust 的 `config/types.rs`）中的实际默认值（512 MB）不一致。这种不一致曾导致用户在排查负载大小问题时产生困惑（如 issue #6622 和 #10800）。因此，修正文档以反映真实默认值是必要的，以提升文档准确性和用户体验。

实现拆解

- 变更入口：修改 `docs/advanced_features/sgl_model_gateway.md` 文件中的配置参考表格。
- 核心修正：将表格中 `--max-payload-size` 参数的默认值从 `256MB` 改为 `512MB`，确保与代码实现一致。 `markdown | `--max-payload-size` | int | 512MB | Maximum request payload |` 修正后，文档准确反映了网关的实际默认配置。
- 无配套改动：本次变更仅限文档，未涉及任何源代码、测试、配置或部署文件的修改，因此无需额外的测试或部署调整。

评论区精华

Review 过程非常简洁，仅由 whybeyoung 进行了批准（APPROVED），未留下任何评论或争议。这表明该文档修正被普遍认可为必要且正确的，无需深入讨论。

风险与影响

风险分析：

- 回归风险：无，因为仅修改文档，不影响运行时逻辑。
- 兼容性风险：无，文档修正不会改变系统行为或接口。
- 安全与性能风险：无。唯一潜在风险是文档修正可能未覆盖所有相关位置，但根据 PR 描述，仅需修改配置参考表格，且已确认与代码一致，因此风险极低。

影响分析：

- 对用户：修正后，用户在查阅网关配置参考时将获得准确的默认值信息，避免因文档错误导致的配置误解或故障排查困难，提升了文档的可信度。
- 对系统：无任何运行时影响，系统功能和行为保持不变。
- 对团队：维护了文档与代码的一致性，减少了未来类似混淆的可能性，体现了良好的文档维护实践。

关联脉络

本次 PR 是纯粹的文档修正，未直接关联其他历史 PR 或 issue。然而，它解决了由 issue #6622 和 #10800 揭示的文档不一致问题，体现了团队对文档准确性的重视。在近期历史 PR 中，类似文档维护的案例（如 PR #22912 更新拼写检查配置、PR #22903 澄清 CI 日志分析规则）也展示了团队在基础设施和文档方面的持续优化。