

PR #24035 完整报告

sgl-project/sglang

[minor] Remove incorrect note after supporting w4a16 moe for DeepSeek V4

合并时间: 2026-04-30 04:26

原文链接: <http://prhub.com.cn/sgl-project/sglang/pull/24035>

执行摘要

该 PR 移除了 DeepSeek-V4 使用指南中一条关于 H200 GPU 应使用特定 SGLang 检查点的过时提示，因为 w4a16 MoE 已获得正式支持，该限制不再适用。变更仅涉及文档，无代码改动。

功能与动机

PR 标题明确说明：在 DeepSeek V4 支持 w4a16 MoE 后，原文档中提示 H200 部署需使用 `sgl-project` 检查点的 Note 已不准确，移除可以避免误导用户。

实现拆解

唯一变更发生在 `docs_new/cookbook/autoregressive/DeepSeek/DeepSeek-V4.mdx` 文件中：

- 删除了 `### 3.1 Basic Configuration` 小节下的 `<Note>` 块，该 Note 原本写道："For H200 GPU deployments, use the SGLang checkpoint under `sgl-project`, not the default DeepSeek checkpoint."
- 共删除 3 行，无新增内容。

无源码变更。

评论区精华

无技术讨论。仅有 `gemini-code-assist` 机器人的自动评论，表示无进一步反馈。

风险与影响

- 风险：无。纯文档修改，不影响任何运行时行为。
- 影响：极小。消除一个已过时的部署指引，避免用户误用。

关联脉络

该 PR 未关联其他 Issue 或 PR，属于独立的小幅文档修正。