

PR #48802 完整报告

vllm-project/vllm

[Docs] fix error key name

合并时间: 2026-07-16 11:16

原文链接: <http://prhub.com.cn/vllm-project/vllm/pull/48802>

执行摘要

- 一句话: 修复 Kubernetes 部署文档中的错误参数名
- 推荐动作: 该 PR 是常规文档修正, 变更简单明确, 可直接合并。建议在相似文档中全局检查参数名一致性。

功能与动机

Kubernetes 部署文档中 `vllm serve` 命令示例使用了错误的参数名

`--max_num_batched_tokens` (下划线分隔), 但正确的 CLI 参数是

`--max-num-batched-tokens` (中划线分隔)。错误的参数名会导致命令执行失败, 影响用户部署体验。

实现拆解

修改了 `docs/deployment/k8s.md` 文件中的两处命令示例, 将下划线分隔的参数名

`--max_num_batched_tokens` 替换为中划线分隔的正确参数名 `--max-num-batched-tokens`。

关键文件:

- `docs/deployment/k8s.md` (模块 部署文档; 类别 docs; 类型 documentation): 唯一修改的文件, 修复了两个 `vllm serve` 命令示例中的参数名错误。

关键符号: 未识别

关键源码片段

`docs/deployment/k8s.md`

唯一修改的文件, 修复了两个 `vllm serve` 命令示例中的参数名错误。

该示例展示 `vllm serve` 命令的正确参数名 (中划线分隔)

args: [

```
"vllm serve mistralai/Mistral-7B-Instruct-v0.3 --trust-remote-code --enable-chunked-prefill --max-num-batched-tokens 1024"
```

]

另一处类似修正

args: [

```
"vllm serve mistralai/Mistral-7B-v0.3 --port 8000 --trust-remote-code --enable-chunked-prefill --max-num-batched-tokens 1024"
```

评论区精华

无实质性讨论。PR 提交后由 [jeejeelee](#) 直接批准，[claude\[bot\]](#) 因 fork 而自动跳过审查。

- 暂无高价值评论线程

风险与影响

- 风险：变更仅涉及文档字符串替换，不包含任何代码逻辑改动，风险极低。但需确认文档中是否还存在其他类似错误参数名未被修正。
- 影响：直接影响阅读 Kubernetes 部署文档的用户，按照正确参数名配置可避免命令失败。影响范围限于文档内容，无系统或性能影响。
- 风险标记：文档变更

关联脉络

- 暂无明显关联 PR