

PR #41419 完整报告

vllm-project/vllm

Fix typo in log message for indexer cache

合并时间: 2026-05-01 06:02

原文链接: <http://prhub.com.cn/vllm-project/vllm/pull/41419>

执行摘要

- 一句话: 修复 indexer 缓存日志中的拼写错误
- 推荐动作: 可直接合并, 无需额外审查。

功能与动机

日志消息中出现拼写错误 "Lighening", 正确拼写应为 "Lightning"。该 PR 旨在修正此 typo, 使日志更准确。

实现拆解

在文件 `vllm/model_executor/layers/deepseek_v4_attention.py` 的第 1055 行, 将字符串 "Lighening Indexer" 改为 "Lightning Indexer"。除此之外没有其他变更。

关键文件:

- `vllm/model_executor/layers/deepseek_v4_attention.py` (模块 模型执行器; 类别 source ; 类型 core-logic) : 唯一的变更文件, 修复了 `DeepseekV4Indexer.__init__` 中 `logger.info_once` 的字符串。

关键符号: `DeepseekV4Indexer.init`

关键源码片段

`vllm/model_executor/layers/deepseek_v4_attention.py`

唯一的变更文件, 修复了 `DeepseekV4Indexer.__init__` 中 `logger.info_once` 的字符串。

```
# vllm/model_executor/layers/deepseek_v4_attention.py
# 第 1053-1057 行: 修正拼写错误
self.use_fp4_kv = self.vllm_config.attention_config.use_fp4_indexer_cache
logger.info_once(
    "Using %s indexer cache for Lightning Indexer.", # 之前为 "Lighening"
    "MXFP4" if self.use_fp4_kv else "FP8",
)
```

评论区精华

无 review 讨论。

- 暂无高价值评论线程

风险与影响

- 风险：变更极小（仅 1 行字符串），不涉及逻辑改动，无回归风险。
- 影响：影响范围仅限于该日志输出，对用户无功能影响，仅改善日志可读性。
- 风险标记：暂无

关联脉络

- PR #41374 [DSV4] Avoid redundant dtype conversion.: 同一模块（`deepseek_v4_attention.py`）的最近改动，说明该文件处于活跃开发中。