

PR #43664 完整报告

vllm-project/vllm

[Misc][Rocm] Remove redundant `AiterUnifiedAttentionBackend` block size log

合并时间: 2026-05-28 11:19

原文链接: <http://prhub.com.cn/vllm-project/vllm/pull/43664>

执行摘要

- 一句话: 移除 ROCm Aiter 后端冗余日志
- 推荐动作: 可直接合入, 无需关注。

功能与动机

PR body 指出日志是冗余的, 且不应作为 warning 级别 (用户无法对此采取任何行动)。块大小管理系统已处理并记录了 INFO 日志, 因此此日志重复且无用。

实现拆解

在 `vllm/v1/attention/backends/rocm_aiter_unified_attn.py` 文件的 `get_preferred_block_size` 方法中, 删除了调用 `logger.warning_once` 输出冗余警告的代码行。方法现在直接返回 64, 不再产生日志。

关键文件:

- `vllm/v1/attention/backends/rocm_aiter_unified_attn.py` (模块 注意力层; 类别 source; 类型 core-logic): 唯一变更文件, 删除了冗余日志调用。

关键符号: 未识别

关键源码片段

`vllm/v1/attention/backends/rocm_aiter_unified_attn.py`

唯一变更文件, 删除了冗余日志调用。

```
# vllm/v1/attention/backends/rocm_aiter_unified_attn.py

class RocmAiterUnifiedAttentionBackend(RocmAttentionBackend):
    # 删除前: classmethod get_preferred_block_size 中有一行冗余的 logger.warning_once
    # 删除后: 直接返回 64, 不再输出警告
    @classmethod
    def get_preferred_block_size(cls, default_block_size: int) -> int:
        # logger.warning_once("[ROCM_AITER_UNIFIED_ATTEN]: Setting kv cache block size to 64.
        # ") # <-- 被删除
        return 64
```

评论区精华

无实质性讨论，AndreasKaratzas 直接批准。

- 暂无高价值评论线程

风险与影响

- 风险：无风险。仅删除一行无关紧要的日志，不影响任何功能逻辑。
- 影响：对用户无影响；减少 ROCm Aiter Unified Attention 后端启动时的一条无用警告日志。
- 风险标记：暂无

关联脉络

- 暂无明显关联 PR