

PR #25707 完整报告

sgl-project/sglang

Log per-request time stats in a dedicated tail step

合并时间: 2026-05-19 09:16

原文链接: <http://prhub.com.cn/sgl-project/sglang/pull/25707>

执行摘要

- 一句话: 提取时间日志记录为私有方法
- 推荐动作: 这是一个小范围重构, 属于代码清理, 值得合并但不需精读。

功能与动机

从 PR body 可知, 提取辅助方法是为了降低 `_stream_output_generation` 的复杂度, 避免将 rank-zero 和时间统计配置等逻辑耦合在累加器 (`_GenerationStreamAccumulator`) 中, 同时作为重构链的一环 (refactor chain ID: extract-maybe-log-time-stats)。

实现拆解

1. 在 `_stream_output_generation` 方法中, 将之前内联的条件判断和 `req.log_time_stats()` 调用替换为 `self._maybe_log_time_stats(req=req)`。
2. 新增 `_maybe_log_time_stats(self, *, req: Req) -> None` 私有方法, 内部保留原始的条件逻辑 (`req.finished()`、`self.ps.attn_tp_rank == 0`、`self.server_args.enable_request_time_stats_logging`)。
3. 该方法位于 `SchedulerOutputStreamer` 类中而非 `_GenerationStreamAccumulator`, 因为条件涉及流管理器的属性 (`self.ps` 和 `self.server_args`), 累加器不应感知这些。
4. `to_payload` 仍为 `NotImplementedError` 存根, 等待后续 PR 接入。

关键文件:

- `python/sglang/srt/managers/scheduler_components/output_streamer.py` (模块 调度器; 类别 source; 类型 core-logic; 符号 `_maybe_log_time_stats`): 唯一变更文件, 提取 `_maybe_log_time_stats` 私有方法并替换调用点。

关键符号: `_maybe_log_time_stats`

关键源码片段

`python/sglang/srt/managers/scheduler_components/output_streamer.py`

唯一变更文件, 提取 `_maybe_log_time_stats` 私有方法并替换调用点。

```
# 新增的 _maybe_log_time_stats 方法, 内部持有原始条件逻辑
def _maybe_log_time_stats(self, *, req: Req) -> None:
    # 仅在请求完成、attn_tp_rank == 0 且配置启用时才落日志
```

```
if (
    req.finished()
    and self.ps.attn_tp_rank == 0
    and self.server_args.enable_request_time_stats_logging
):
    req.log_time_stats()

# 调用点替换: for 循环中
for req in reqs:
    if req is skip_req:
        continue
    if req.finished() and req.finished_output:
        continue
    acc.accept(req=req)
    self._maybe_log_time_stats(req=req) # 由原来 10 行内联替换为单行
```

评论区精华

无 review 讨论，PR 由作者自行合并。

- 暂无高价值评论线程

风险与影响

- 风险：变更范围极小（16 行 diff），逻辑完全等价，不引入新功能或性能风险。唯一潜在风险是 `_maybe_log_time_stats` 被其他代码意外调用（但为私有方法），或重构链后续变更与本变更冲突。
- 影响：仅限于 `_stream_output_generation` 方法内部，无外部行为变化。影响面小，仅对后续重构链有积极影响。
- 风险标记：暂无

关联脉络

- 暂无明显关联 PR