

PR #25438 完整报告

sgl-project/sglang

Convert forward_pass_device_timer to None-init

合并时间: 2026-05-16 09:20

原文链接: <http://prhub.com.cn/sgl-project/sglang/pull/25438>

执行摘要

- 一句话: 将 `forward_pass_device_timer` 初始化为 `None`
- 推荐动作: 可以合并。这是一个小型机械重构, 提升了代码可维护性, 无功能变化。

功能与动机

PR body 指出 `forward_pass_device_timer` 之前仅在启用设备定时器指标时条件赋值, 导致读取点必须用 `hasattr` 防御属性不存在的情况。将其无条件初始化为 `None` 并用 `is None` 守卫, 使真正的不变性 (启用设备定时器指标时有值, 否则为 `None`) 直接体现在类型中。

实现拆解

1. 在 `SchedulerMetricsMixin.__init__` 中, 处于 `if ENABLE_METRICS_DEVICE_TIMER:` 条件赋值之前, 添加一行 `self.forward_pass_device_timer: Optional[DeviceTimer] = None`, 确保属性始终存在。
2. 在 `install_device_timer_on_runners` 方法中, 将守卫条件从 `if not hasattr(self, "forward_pass_device_timer"): return` 改为 `if self.forward_pass_device_timer is None: return`。
3. 在 `_init_fpm` 方法中, 将读取点的 `if hasattr(self, "forward_pass_device_timer"):` 改为 `if self.forward_pass_device_timer is not None:。`

关键文件:

- `python/sglang/srt/observability/scheduler_metrics_mixin.py` (模块 调度指标; 类别 source; 类型 core-logic; 符号 `init`, `install_device_timer_on_runners`, `_init_fpm`): 唯一变更文件, 包含条件初始化添加和两处守卫替换, 是整个 PR 的核心。

关键符号: `init`, `install_device_timer_on_runners`, `_init_fpm`

关键源码片段

`python/sglang/srt/observability/scheduler_metrics_mixin.py`

唯一变更文件, 包含条件初始化添加和两处守卫替换, 是整个 PR 的核心。

```
# python/sglang/srt/observability/scheduler_metrics_mixin.py (head)
```

```
class SchedulerMetricsMixin:
```

```

def init_metrics(self, ...):
    # ... 前面的初始化代码 ...

    self.fwd_occupancy = float("nan")

    # [ 新增 ] 无条件初始化为 None, 确保属性始终存在
    self.forward_pass_device_timer: Optional[DeviceTimer] = None

    if ENABLE_METRICS_DEVICE_TIMER:
        self._device_timer_window_batch_count = 0
        self._device_timer_window_gpu_time = 0.0
        self._device_timer_window_start = None

        def _wrap_execution_reporter(**kwargs):
            self._device_timer_window_gpu_time += kwargs["t"]
            if self.enable_metrics:
                self.metrics_collector.increment_forward_execution_seconds(**kwargs)

        self.forward_pass_device_timer = DeviceTimer(
            reporter=_wrap_execution_reporter,
        )

    # ... 剩余初始化代码 ...

def install_device_timer_on_runners(self: Scheduler):
    # [ 修改 ] 用 is None 替代 hasattr, 更直接表达语义
    if self.forward_pass_device_timer is None:
        return
    timer = self.forward_pass_device_timer
    self.tp_worker.model_runner.device_timer = timer
    # ... 注入到 draft worker 的代码 ...

def _init_fpm(self: Scheduler):
    # ...
    def _fpm_device_timer_reporter(t, **kwargs):
        self._fpm_gpu_time_acc += t

    # [ 修改 ] 同样用 is not None 替代 hasattr
    if self.forward_pass_device_timer is not None:
        self.forward_pass_device_timer.add_reporter(_fpm_device_timer_reporter)
    else:
        self.forward_pass_device_timer = DeviceTimer(
            reporter=_fpm_device_timer_reporter,
        )
    self._fpm_uses_device_timer = True
    # ...

```

评论区精华

本 PR 无 review 讨论。

- 暂无高价值评论线程

风险与影响

- 风险：风险极低。变更仅涉及两个守卫条件和一个无条件初始化，逻辑等价。如果 `forward_pass_device_timer` 在 `__init__` 外被删除或重写，可能导致 `is None` 守卫无法正确处理，但这种用法在代码库中不存在。未引入新的 `import` 或外部依赖。
- 影响：直接影响 `SchedulerMetricsMixin` 及其子类，以及 `install_device_timer_on_runners` 的调用方。对系统行为无影响，是纯粹的重构。
- 风险标记：暂无

关联脉络

- PR #25444 Bundle Scheduler rank/size fields into a frozen ParallelState: 同属调度器观测性重构链，均为提升代码清晰度的机械式重构。
- PR #25445 Inject ParallelState into ProfilerV2: 同为调度器观测性模块的重构，且 PR body 中 Refactor chain ID 表明属于同一重构链。
- PR #25442 Lift forward_ct/cur_batch and use direct access in watchdog: 同样是移除防御性 `getattr` 的模式，涉及同一文件或相近模块。