

# PR #25446 完整报告

sgl-project/sglang

Fix V2 trace filename collisions when DP/PP/EP enabled

合并时间: 2026-05-16 09:24

原文链接: <http://prhub.com.cn/sgl-project/sglang/pull/25446>

## 执行摘要

- 一句话: 修复多并行策略 trace 文件覆盖 bug
- 推荐动作: 建议合并。这是一个明确的 bugfix, 改动微小且正确性高。

## 功能与动机

在启用 DP (数据并行)、PP (流水线并行) 或 EP (专家并行) 时, `_ProfilerTorch` 生成的 trace 文件因文件名缺少区分性后缀而被多 rank 覆盖, 只剩最后一个 rank 的数据, 导致其他 rank 的性能数据丢失。PR body 明确指出 'every rank wrote to a filename like `<id>-TP-<n>.trace.json` with no rank-disambiguating suffix — so multi-rank runs overwrote each other's traces, losing all but one rank's profile.'

## 实现拆解

仅修改一个文件 `python/sglang/srt/utils/profile_utils.py`, 对 `_ProfilerTorch.stop()` 方法中的 6 处 `getattr` 调用进行替换:

1. 将 `if getattr(self, "dp_size", 1) > 1:` 改为 `if self.ps.dp_size > 1:`, `filename_parts.append(f"DP-{{getattr(self, 'dp_rank', 0}}")` 改为 `filename_parts.append(f"DP-{{self.ps.dp_rank}}")`。
2. 相同模式对 `pp_size/pp_rank` 和 `moe_ep_size/moe_ep_rank` 各两处进行修改。
3. 关键点: `self.ps` 是此前 `inject-parallel-state-profiler` 重构中注入的 `ParallelState` 对象, 其保存了正确的并行配置。
4. 无测试、配置或部署配套改动。

关键文件:

- `python/sglang/srt/utils/profile_utils.py` (模块 性能分析; 类别 source; 类型 core-logic) : bug 定位和修复的全部代码都在此文件, trace 文件名生成逻辑的 6 行 `getattr` 替换为 `self.ps` 直接访问。

关键符号: 未识别

## 评论区精华

无 reviewer 评论。仅有一条 `gemini-code-assist` 的 quota 提示, 不相关。

- 暂无高价值评论线程

## 风险与影响

- 风险：低风险。如 `self.ps` 不存在，会直接抛出 `AttributeError` 而非静默回退为 `size=1`，但 `inject-parallel-state-profiler` 重构已确保 `self.ps` 可用。此外，`self.ps.pp_rank` 等字段的命名与之前的 `getattr` 默认值字段名一致，不存在兼容性问题。
- 影响：影响范围仅限于启用 DP、PP 或 EP 时的性能分析场景。修复后 `trace` 文件将正确包含 `DP-{rank}`、`PP-{rank}`、`EP-{rank}` 后缀，所有 `rank` 的 `trace` 被保留。对未启用这些并行策略的用户无行为变化。
- 风险标记：低风险修复

## 关联脉络

- PR #25449 Convert discarded-value ternary to a plain if statement: 与本次改动在同一重构链 (`tom_pr_chain/tom_refactor_202605a`) 中，都是对 `profile_utils` 相关代码的清理与简化。