

PR #27405 完整报告

sgl-project/sglang

Don't write crash dump on graceful exit

合并时间: 2026-06-06 08:06

原文链接: <http://prhub.com.cn/sgl-project/sglang/pull/27405>

执行摘要

- 一句话: 修复优雅退出时误写 crash dump 的问题
- 推荐动作: 本次变更为一次小范围、低风险的 bugfix, 逻辑清晰, 改动量小。建议快速合并。

功能与动机

每次优雅关闭都会产生 crash dump 文件, 容易误导运维人员认为发生了崩溃, 且不必要的磁盘写入影响使用体验。PR 描述指出: "A graceful exit is not a crash", 因此跳过 dump 是合理的修复。

实现拆解

1. 删除崩溃转储调用: 在 `python/sglang/srt/managers/tokenizer_manager.py` 的 `sigterm_watchdog` 方法中, 当请求全部排空 (`remain_num_req == 0`) 时, 删除对 `self.dump_requests_before_crash()` 的调用, 直接 `break` 退出循环。
2. 保持其他逻辑不变: Unhealthy 状态和强制关闭分支 (`SGL_FORCE_SHUTDOWN`) 仍然保留原有行为, 仅在排空后的清理路径中移除 dump 调用。

关键文件:

- `python/sglang/srt/managers/tokenizer_manager.py` (模块 请求管理; 类别 source; 类型 core-logic): 包含核心方法 `sigterm_watchdog`, 本次变更的唯一修改文件。

关键符号: `sigterm_watchdog`

关键源码片段

`python/sglang/srt/managers/tokenizer_manager.py`

包含核心方法 `sigterm_watchdog`, 本次变更的唯一修改文件。

```
# 位于 TokenizerManager 类中
async def sigterm_watchdog(self):
    """Gracefully drain requests on SIGTERM, then exit."""
    while not self.gracefully_exit:
        await asyncio.sleep(5)

# Drain requests
while True:
```

```

remain_num_req = len(self.rid_to_state)
remaining_rids = list(self.rid_to_state.keys())

if self.server_status == ServerStatus.UnHealthy:
    # 健康检查失败时强制退出，保留 crash dump
    logger.error(
        "Signal SIGTERM received while health check failed. Force exiting."
    )
    self.dump_requests_before_crash()
    self.force_exit_handler()
    break

elif get_bool_env_var("SGL_FORCE_SHUTDOWN"):
    # 强制关闭标志，直接退出，不 dump
    logger.error(
        "Signal SIGTERM received while force shutdown flag set. Force exiting."
    )
    self.force_exit_handler()
    break

logger.info(
    f"Gracefully exiting... Remaining number of requests {remain_num_req}. "
    f"Remaining requests {remaining_rids=}."
)
if remain_num_req > 0:
    await asyncio.sleep(5)
else:
    # 所有请求已排空：优雅退出，不应写入 crash dump
    break

kill_process_tree(os.getpid(), include_parent=True)
sys.exit(0)

```

评论区精华

该 PR 没有 review 评论，仅有一条来自 `gemini-code-assist[bot]` 的自动提示（每日配额已满），无实质性讨论。

- 暂无高价值评论线程

风险与影响

- 风险：风险极低。该变更仅在优雅退出的排空路径中移除一行调用，与 crash dump 无关的流程不受影响。Unhealthy 状态和强制关闭分支保留了 crash dump 功能，不会丢失真正的崩溃信息。
- 影响：用户侧：优雅关闭后不再产生无意义的 crash dump 文件，日志更清晰。运维侧：减少了误导性的报警。系统侧：无性能影响。影响范围仅限于 `sigterm_watchdog` 方法的一个分支，非常局部。

- 风险标记：变更极小，风险低

关联脉络

- 暂无明显关联 PR