PR #32910 完整报告

sgl-project/sglang

[DeepSeek-V4] Fix nvcc 13 crash building the topk_v2 kernel

合并时间：2026-08-03 10:59

原文链接：http://prhub.com.cn/sgl-project/sglang/pull/32910

PR #32910 分析：修复 nvcc 13 编译 topk_v2 JIT 内核崩溃

执行摘要

本 PR 以 7 行改动修复了 CUDA 13.x 工具链下 cicc 编译器构建 dpsk_v4_topk_v2 JIT 内核时的段错误（进程以 exit 139 退出）问题。该问题导致 DeepSeek-V4 与 GLM DSA 模型在 Hopper / Blackwell 主机上启动时于 CUDA graph capture 阶段即崩溃（issue #32830）。修复方式是将 DSMEM 映射指针放入按值传递的 TopKProblem 副本中，隔离出 problem_transform epilogue 的加载路径，运行语义不变。PR 曾附带的 sm_90a 编译回归测试因 corner case 价值争议在 review 后移除，最终合并内容仅含一个文件的源码修复。

功能与动机

- issue #32830 报告：用户在 4 卡 H100 上以 CUDA 13.2 启动 DeepSeek-V4-Flash，服务在 nvcc 编译内核时以段错误终止。
- PR body 的定位结论：这是 cicc 优化器崩溃而非源码错误——topk_small_batch_kernel 将 cluster.map_shared_rank(topk_indices, worker_rank) 的 DSMEM 地址写入 problem.out，而 problem_transform epilogue 经同一指针加载；作者在 sglang 外以报告者工具链（nvcc 13.2.86 + gcc 13.4）对 load_jit() 生成的精确翻译单元复现 exit 139，并二分定位到 "DSMEM 指针 + 后续 load" 的组合触发，-Xcicc -O1 可编译未修补文件，锁定为优化器缺陷。
- 波及面不止 DSV4：issue 评论区确认 GLM-5.2 在 CUDA 13.2 下同样启动失败，且 SGLANG_OPT_USE_TOPK_V2=0 可绕过；topk_v2.cuh 是 DSV3.2 / DSV4 与 GLM DSA 共享的 top-k 变换内核，因此本修复具有跨模型收益。

实现拆解

1. 根因定位：在 sglang 外独立复现崩溃，得到与报告一致的警告与 exit 139；二分确认崩溃来自 DSMEM 指针与后续 load 的组合，-Xcicc -O1 可编译证明是 cicc 优化器 bug，而非源码错误。
2. 核心修复：python/sglang/kernels/jit/csrc/deepseek_v4/topk_v2.cuh 中 topk_small_batch_kernel 的 cluster 分支（blockIdx.y != worker_rank 侧），把 " 直接对 problem.out 赋映射地址再调 Cluster::forward" 改为：先浅拷贝 peer_problem = problem，只对 peer_problem.out 赋映射地址，再按值传入 Cluster::forward。因 Cluster::forward 本就按值接收 TopKProblem，被选中的 rank 仍通过自己的 topk_indices 读回同一份字节，与相邻的 Register4 / Streaming 分支语义一致，无行为变

化。

3. 兼容性验证：补丁后同一翻译单元在 nvcc 13.2.86 与 12.6 下均编译通过；无 GPU 运行验证（作者仅有 A40 / sm_86，该内核在该架构下不支持 cluster_dims），244 个既有正确性用例未能在本地执行，PR body 明确请求在 Hopper / Blackwell 上补跑。
4. 测试配套演进：初始提交新增 test_topk_v2_compiles_for_sm90a（用 nvcc -ptx -arch=sm_90a 编译翻译单元，无需集群 GPU，缺 nvcc 时跳过），经 DarkSharpness review 认为 corner case 不值得保留专门回归测试后删除，test_topk_v2.py 恢复 PR 前内容；/rerun-test test_topk_v2.py 在 1-gpu-h100 上通过，CI 红色均为基础设施 503 故障的 fast-fail 级联。

python/sglang/kernels/jit/csrc/deepseek_v4/topk_v2.cuh

唯一变更文件：topk_small_batch_kernel 的 cluster 分支将 DSMEM 映射指针移入 TopKProblem 副本，规避 CUDA 13.x cicc 优化器段错误，解除 DSV4 / GLM DSA 在 Hopper / Blackwell 上的启动崩溃。

```
// 整理自 python/sglang/kernels/jit/csrc/deepseek_v4/topk_v2.cuh 中的
// topk_small_batch_kernel 函数，展示其 cluster 协作分支（修复后版本）。
// 前置逻辑已省略，包括 smem 初始化、problem 填充、worker_rank 计算等。
if (blockIdx.y == worker_rank) {
    // 本 rank 直接走 streaming 分支，读取自己的 topk_indices
    Streaming::forward<kPDL>(problem, &smem);
} else {
    auto cluster = cooperative_groups::this_cluster();

    // 修复说明：map_shared_rank 返回 shared::cluster（DSMEM）的映射地址。
    // 旧代码把该地址直接写入 problem.out，而 problem_transform epilogue
    // 会经由同一指针加载数据；cicc 在 CUDA 13.x 上编译这一模式时发生
    // 段错误（exit 139），导致 dpsk_v4_topk_v2 JIT 模块构建失败，DSV4
    // 与 GLM DSA 服务在 Hopper / Blackwell + CUDA 13 环境启动即崩溃
    // （issue #32830）。修复方式是把映射别名放进 TopKProblem 的浅拷贝
    // peer_problem，仅将其传给按值接收参数的 Cluster::forward；被选中的
    // rank（blockIdx.y == worker_rank）仍通过自己的 topk_indices 读回
    // 同一份字节，运行语义完全不变。
    auto peer_problem = problem;
    peer_problem.out = cluster.map_shared_rank(topk_indices, worker_rank);
    Cluster::forward<kPDL>(peer_problem, &smem); // 写入 peer 的输出共享内存
    cluster.sync();
}
```

评论区精华

- DarkSharpness 对新增编译测试的 diff 留言："I don't think we should include this test for this extremely corner case. What do you think @BBuf @zcnrex"——作者随后在 issue 评论中确认删除该测试；DarkSharpness 最终 APPROVED 并评价 "LGTM. nvcc crash is really sick 😱"。
- 用户 lucashaha 确认 GLM-5.2 同样受影响，并实测 SGLANG_OPT_USE_TOPK_V2=0 可绕过启动崩溃；作者回应 topk_v2.cuh 为 DSV3.2 / DSV4 与 GLM DSA 共享内核，修复落

地后应无需环境变量即可正常构建，若仍失败可再 ping 汇报。

- 作者澄清红色 CI 为基础设施故障：sglang-kernel 0.4.5+cu130 wheel 下载 503、actions/checkout 同步 503，11 个 red job 均为 fast-fail 级联，未运行任何测试；/rerun-test test_topk_v2.py 在同一 commit 上返回绿色。

风险与影响

- 风险：一是 244 个正确性用例未在 Hopper / Blackwell 上运行，虽逻辑等价论证充分，但 cluster 协作路径与 DSMEM 读写仍需目标架构确认；二是回归测试被移除后该 cicc 崩溃无自动防护，未来重构可能重新引入；三是 JIT 树未对 CUDA 13.2 全量编译，启动路径上仍可能存在其他 cicc 崩溃（但无其他内核使用 map_shared_rank，此触发器唯一）；四是修复正确性依赖 Cluster::forward 按值传参的既有约定。
- 影响：对 CUDA 13.x + Hopper / Blackwell 用户是启动即崩的阻断性问题的解除，直接受益模型为 DeepSeek-V4 与 GLM DSA 系列；对 CUDA 12.x 用户与已用环境变量绕过的用户无行为变化。改动仅限单文件 7 行，运行时行为不变，风险面小。团队层面沉淀了一套可复用的编译器崩溃诊断范式（独立翻译单元复现 + 二分 + -Xcicc -O1 缓解实验），对后续 CUDA 13 全面升级有参考价值。

关联脉络

- 直接对应 issue #32830；issue 评论将影响面扩展到 GLM-5.2，与 PR #33100（GLM 5.2 CP-v2 修复）同属 GLM 型号维护线，共享 DSA / topk_v2 内核路径。
- 与近期 DeepSeek-V4 系列 bugfix（PR #33276 DSpark 专家加载 scale 跳过、PR #31727 gfx950 FP8 scale 错位）共同显示 DSV4 在新工具链、新硬件、新量化路径上的兼容性修复正在收敛，为 CUDA 13 全面升级与 Blackwell 支撑铺路。
- 本次修复仅触及 JIT 内核的 cluster 通信路径，未进入调度 / 显存管理核心，与 DSPARK 采样支持（PR #33298）、HiCache L2（PR #33112）等 DeepSeek 技术栈的演进相互独立。