

PR #27163 完整报告

sgl-project/sglang

[AMD] Disable AITER custom all-gather in DeepSeek-R1-MXFP4 8-GPU test

合并时间: 2026-06-04 04:38

原文链接: <http://prhub.com.cn/sgl-project/sglang/pull/27163>

执行摘要

- 一句话: 禁用在 MI35x TP=8 上导致 crash 的 AITER all-gather
- 推荐动作: 值得关注的是 PR 中对根因的详细分析 (AITER CUDA-graph 缓冲区注册偏移问题), 这对理解多 GPU 通信与 CUDA-graph 交互有帮助。生产环境下建议跟踪 AITER 的修复, 及时恢复自定义 all-gather 以保持性能。

功能与动机

MI35x 8-GPU 上运行 DeepSeek-R1-MXFP4 测试时, decode 阶段出现多 rank 的 "Memory access fault by GPU node-N ... Reason: Unknown" 错误, 而权重加载和 CUDA-graph 捕获均正常。二分定位到 PR25093 引入的 AITER 自定义 all-gather 是回归点。

实现拆解

1. 定位问题: 在 `test/registered/amd/test_deepseek_r1_mxfp4_8gpu.py` 中, 两个测试类 (`TestDeepseekR1MXFP4` 和 `TestDeepseekR1MXFP4MTP`) 的 `setUpClass` 方法中添加 `envs.SGLANG_USE_AITER_AG.set(False)`, 禁用 AITER 自定义 all-gather。
2. 注释说明: 为每个改动处添加注释, 解释这是临时解决方案, 并指向 PR 描述以了解根因。
3. 验证测试: 在禁用该标志后, 4 项测试 (GSM8K 精度、单 batch 速度、MTP GSM8K 精度、MTP 单 batch 速度) 均通过。

关键文件:

- `test/registered/amd/test_deepseek_r1_mxfp4_8gpu.py` (模块 AMD 测试; 类别 test; 类型 test-coverage): 添加 `SGLANG_USE_AITER_AG=False` 环境变量以绕过 AITER all-gather bug, 所有 CI 测试由此通过。

关键符号: 未识别

关键源码片段

`test/registered/amd/test_deepseek_r1_mxfp4_8gpu.py`

添加 `SGLANG_USE_AITER_AG=False` 环境变量以绕过 AITER all-gather bug, 所有 CI 测试由此通过。

```
# test/registered/amd/test_deepseek_r1_mxfp4_8gpu.py
```

```

class TestDeepseekR1MXFP4(CustomTestCase):
    @classmethod
    def setUpClass(cls):
        cls.model = DEEPSEEK_R1_MODEL_PATH
        cls.base_url = DEFAULT_URL_FOR_TEST

        # Workaround: AITER custom all-gather corrupts CUDA-graph IPC buffer
        # registration and triggers a decode-time "Memory access fault" on
        # MI35x TP=8. Disable until the AITER-side fix lands (see PR body).
        envs.SGLANG_USE_AITER_AG.set(False)

        other_args = [
            "--tp", "8",
            "--chunked-prefill-size", "131072",
            "--model-loader-extra-config", '{"enable_multithread_load": true}',
        ]
        cls.process = popen_launch_server(
            cls.model, cls.base_url,
            timeout=SERVER_LAUNCH_TIMEOUT, other_args=other_args,
        )

class TestDeepseekR1MXFP4MTP(CustomTestCase):
    @classmethod
    def setUpClass(cls):
        cls.model = DEEPSEEK_R1_MODEL_PATH
        cls.base_url = DEFAULT_URL_FOR_TEST

        envs.SGLANG_ENABLE_OVERLAP_PLAN_STREAM.set(True)
        # Same AITER custom all-gather workaround as TestDeepseekR1MXFP4 above.
        envs.SGLANG_USE_AITER_AG.set(False)

        other_args = [
            "--tp", "8",
            "--chunked-prefill-size", "131072",
            "--speculative-algorithm", "EAGLE",
            "--speculative-num-steps", "3",
            "--speculative-eagle-topk", "1",
            "--speculative-num-draft-tokens", "4",
            "--model-loader-extra-config", '{"enable_multithread_load": true}',
        ]
        cls.process = popen_launch_server(
            cls.model, cls.base_url,
            timeout=SERVER_LAUNCH_TIMEOUT, other_args=other_args,
        )

```

评论区精华

无实质性讨论。PR 获得 HaiShaw 批准，没有 review 评论。

- 暂无高价值评论线程

风险与影响

- 风险：该 PR 本身风险很低，仅修改测试配置。但需注意：
 - 禁用 AITER custom all-gather 后，DeepSeek-R1-MXFP4 在 MI35x 上的 all-gather 性能可能下降，因为 fallback 到原生实现（如 NCCL）。
 - 一旦 AITER 上游修复问题，需要删除此工作区并重新启用标志。
 - 影响：直接影响 SGLang 在 AMD MI35x 平台上的 DeepSeek-R1-MXFP4 测试流程：修复了 8-GPU 配置下 decode 阶段的内存访问故障，使 4 项关键测试（GSM8K 精度、吞吐量、MTP 版精度与吞吐量）恢复正常。对用户无直接影响，因为测试环境变量仅在 CI 中设置。
 - 风险标记：临时工作区，依赖上游修复

关联脉络

- PR #25093 [AMD] Enable AITER custom all-gather on ROCm: 引入 AITER custom all-gather 并默认启用，导致本 PR 需要禁用它来绕过 bug。