PR #25209 完整报告

sgl-project/sglang

[AMD] Add amd jit clamp position bench ci

合并时间：2026-05-14 15:03

原文链接：http://prhub.com.cn/sgl-project/sglang/pull/25209

执行摘要

- 一句话：新增 AMD CI benchmark 注册
- 推荐动作：建议合入。属于标准 CI 接入变更，代码简洁，已验证 AMD 环境通过。

功能与动机

为 AMD 设备补全 JIT kernel 的 CI benchmark，之前已有 CUDA 注册，现在需要 AMD 版本。

实现拆解

在 python/sglang/jit_kernel/benchmark/bench_clamp_position.py 中：

1. 导入 register_amd_ci 函数：将 import 行改为 from sglang.test.ci.ci_register import register_amd_ci, register_cuda_ci。
2. 调用 register_amd_ci：新增 register_amd_ci(est_time=16, suite="jit-kernel-unit-test-amd")，并与原有的 register_cuda_ci 并列。

关键文件：

- python/sglang/jit_kernel/benchmark/bench_clamp_position.py（模块 JIT 核；类别 source；类型 dependency-wiring）：单文件变更，添加 AMD CI 注册，使 benchmark 可在 AMD 设备上自动执行。

关键符号：未识别

关键源码片段

python/sglang/jit_kernel/benchmark/bench_clamp_position.py

单文件变更，添加 AMD CI 注册，使 benchmark 可在 AMD 设备上自动执行。

```
# bench_clamp_position.py (head)
import itertools
import torch
import triton
import triton.testing

from sglang.jit_kernel.benchmark.utils import (
    DEFAULT_DEVICE,
    get_benchmark_range,
```

```
    run_benchmark,
)
from sglang.jit_kernel.clamp_position import clamp_position_cuda
from sglang.srt.utils import get_compiler_backend
# 同时导入 CUDA 和 AMD 的 CI 注册函数
from sglang.test.ci.ci_register import register_amd_ci, register_cuda_ci

# 注册 CUDA CI benchmark
register_cuda_ci(est_time=13, suite="stage-b-kernel-benchmark-1-gpu-large")
# 新增 AMD CI benchmark 注册，预估耗时 16 秒，归入 jit-kernel-unit-test-amd 套件
register_amd_ci(est_time=16, suite="jit-kernel-unit-test-amd")

SIZE_LIST = get_benchmark_range(...)
# ... 后续 benchmark 逻辑不变
```

评论区精华

无 review 评论。

- 暂无高价值评论线程

风险与影响

- 风险：风险很低。仅添加一行注册调用，不影响原逻辑；若 AMD CI 环境缺失依赖可能导致 benchmark 失败，但仅影响 CI 流程。
- 影响：对用户无直接影响。对 AMD CI 管道增加一个 benchmark 任务（估时 16 秒），提升 JIT kernel 在 AMD 平台上的可观测性。
- 风险标记：暂无

关联脉络

- PR #25210 [AMD] Add amd jit resolve token ids bench ci: 类似动机：为另一个 JIT kernel benchmark 添加 AMD CI 注册，属于同一批次基础设施搭建。