

PR #34147 完整报告

sgl-project/sglang

[AMD] [CI] Register the DeepSeek-V4-Pro-DSpark MI35x nightly job so its suite actually runs

合并时间: 2026-08-10 05:10

原文链接: <http://prhub.com.cn/sgl-project/sglang/pull/34147>

执行摘要

- 一句话: 修复 AMD 夜测中 DSV4-DSpark 套件从不运行
- 推荐动作: 值得关注, 但不需要逐行精读。核心价值在于: 1) 演示了 `register_amd_ci` 只做注册、必须由 workflow job 调用 `run_suite.py --suite` 才能执行这一机制, 避免“注册即沉睡”的坑; 2) 展示新增 AMD nightly job 的标准接线 (机器池、容器镜像、依赖安装、`job_select`、`check-all-jobs`、`timeout` 预算); 3) body 中“为什么只跑单 backend 腿”与“为什么 `timeout` 取 7200”的论证是值得借鉴的 CI 设计权衡。若你维护 AMD / nightly CI 或 DeepSeek-V4 相关测试, 建议阅读; 普通功能开发者可跳过。

功能与动机

30964 添加了 `test/registered/amd/test_deepseek_v4_pro_fp4_dspark.py`, 这是唯一覆盖 AMD DSpark unified-KV 路径的测试, 注册时只写了 `register_amd_ci(..., suite='nightly-amd-8-gpu-mi35x-deepseek-v4-pro-dspark', nightly=True)`。PR body 指出: No job in `nightly-test-amd.yml` or `nightly-test-amd-rocm720.yml` invokes that suite name, so the test never runs; `nightly=True` 只是让它不进 PR CI, 并不会被任何地方调度, `git grep` 在 `main` 上只能匹配到测试文件自身。而 #30964 的合并评审明确要求该 suite 在 MI35x 上以 `unified_kv_triton` 运行并确认 GSM8K 精度与 DSpark acceptance, 此前只能手工复现。

实现拆解

1. 定位 dormant suite: 在 `main` 上执行 `git grep nightly-amd-8-gpu-mi35x-deepseek-v4-pro-dspark`, 确认该 suite 名只出现在 #30964 的测试文件注册里, 两个 AMD nightly workflow 都没有 job 调用它; `register_amd_ci` 只负责把测试文件关联到 suite 名, 执行必须由某个 job 调用 `run_suite.py --suite <name>` 完成。

2. 新增 nightly job: 在 `.github/workflows/nightly-test-amd-rocm720.yml` 的 `DeepSeek-V4 MI35x` 分组中增加 `nightly-8-gpu-mi35x-deepseek-v4-pro-dspark-rocm720`, 构建方式对齐相邻 `nightly-8-gpu-mi35x-deepseek-v4-pro-rocm720`: `runs-on: linux-mi35x-gpu-8`、先跑 `scripts/ci/amd/ensure_vram_clear.sh` rocm、使用 ROCm 7.2 容器 (`ENABLE_CACHE_HOST=1`)、依赖安装带 `--skip-test-time-deps` (GSM8K 不需要 `lmms-eval / human-eval`) 并补 `pip install tabulate` 支撑结果表格输出。
3. 运行套件与环境变量: 执行 `python3 run_suite.py --hw amd --suite nightly-amd-8-gpu-mi35x-deepseek-v4-pro-dspark --nightly --timeout-per-file 7200`; 通过 `-e` 注入 `SGLANG_HACK_FLASHMLA_BACKEND=unified_kv_triton` (锁定被测 backend) 与 `SGLANG_MOE_COPY_WEIGHT_VIEWS_BEFORE_H2D=1`, 并把 `GITHUB_STEP_SUMMARY` 指向容器外路径, 使 GSM8K 精度与 DSpark 接受长度进入 job summary。与相邻 V4 job 刻意不同: 只跑 `unified_kv_triton` 单腿、不跑 `triton` 腿, 因为测试目标就是 unified-KV 的 `target-hidden` 注入路径。
4. 接入调度与失败联动: 将 job 名追加到 workflow 的 `job_select` 下拉选项, 使其可被单独 dispatch; 同时加入 `check-all-jobs.needs`, 保证该 job 失败会让整个 nightly 标红, 而不是悄悄缺席。
5. 验证与时间预算: 作者在分支上 dispatch 一次 (run 31291406552, 其余 46 个 job 跳过), 被测代码即已合入 #30964 的 main; 套件 1/1 通过、3 个测试耗时 2906.13s, GSM8K 精度 0.9515 (门禁 > 0.92)、DSpark 平均接受长度 3.730 (门禁 > 3.0)。实测排队 51 min、job 墙钟 55m23s, 套件实际约 2930s 而注册的 `est_time=7200` 是其 2.5 倍, PR body 明确建议后续收紧以利于 nightly 分区平衡。

本 PR 自身不新增测试文件, 也不改模型、内核或调度代码; 属于纯 CI 配置配套。

关键文件:

- `.github/workflows/nightly-test-amd-rocm720.yml` (模块 夜测配置; 类别 infra; 类型 configuration; 符号 `nightly-8-gpu-mi35x-deepseek-v4-pro-dspark-rocm720`, `nightly-amd-8-gpu-mi35x-deepseek-v4-pro-dspark`): 唯一变更文件: 新增 `nightly-8-gpu-mi35x-deepseek-v4-pro-dspark-rocm720` job 并把 suite 名接入 `job_select` 与 `check-all-jobs`, 是本 PR 修复“测试沉睡”的全部内容。

关键符号: `nightly-8-gpu-mi35x-deepseek-v4-pro-dspark-rocm720` (GitHub Actions job), `run_suite.py` (CI 套件执行入口)

评论区精华

该 PR 没有任何 review 评论 (comments 与 review comments 均为 0), HaiShaw 直接 APPROVED。可提取的设计讨论都沉淀在 PR body 中, 两处权衡值得记录:

- 为什么只跑 `unified_kv_triton` 单腿: 相邻 V4-Pro / V4-Flash job 都有 `unified_kv_triton` + `triton` 双腿, 但 DSpark 测试默认 backend 就是前者, `triton` 腿不会触碰它要验证的 `unified-KV target-hidden` 注入路径, 属于“照抄相邻 job 的路径依赖”需要被主动打破。
- 为什么 `timeout-minutes: 300`: `SERVER_LAUNCH_TIMEOUT` 本身 5400s 只覆盖 1.6T checkpoint 的加载, 之后还有 1319 题 GSM8K 推理, `--timeout-per-file 7200` 与注册的 `est_time=7200` 一致, 是必要余量而非随意放大。

- 暂无高价值评论线程

风险与影响

- 风险：风险面聚焦在 CI 资源与 job 联动，而非产品代码：
 - 资源占用与排队：新增 job 占用 linux-mi35x-gpu-8 整机约 1 小时墙钟（实测 55m23s），且实测排队曾达 51 min；接入 check-all-jobs.needs 后，排队过久或机器不健康可能让整个 nightly 失败，反而掩盖其他 job 的问题。
 - 超时预算：timeout-minutes: 300 相对实测墙钟仍有余量，但 SERVER_LAUNCH_TIMEOUT 5400s 的 checkpoint 加载若因网络或存储抖动变慢，叠加队列等待可能触顶。
 - 单路径覆盖：该 job 不覆盖 triton 等非 DSpark 路径；未来若 SGLANG_HACK_FLASHMLA_BACKEND 的默认值或命名变化，测试可能失效而不易察觉。
 - est_time 虚高：est_time=7200 相对套件实际约 2930s 保守约 2.5 倍，长期影响 nightly 分区调度器的资源预留精度。
 - 影响：对用户：无直接运行时影响，但对 AMD MI35x 上运行 DeepSeek-V4-Pro-DSpark 的部署增加了一道持续回归防线，GSM8K 精度与投机接受长度会成为每次 nightly 的自动门禁。对系统：AMD nightly 增加一个约 1 小时的 8 卡 MI35x job，并因 check-all-jobs 纳入整体健康度。对团队：确立了“注册 suite 必须配套调度 job”的 CI 纪律；job_select + check-all-jobs 的接线模式可复用到后续新增 nightly 测试。
- 风险标记：仅覆盖 unified_kv_triton 单路径，check-all-jobs 全量联动增加误报面，MI35x 资源排队延长 nightly 时长，est_time 预留 2.5 倍影响分区平衡

关联脉络

- PR #30964 [AMD] Support DeepSeek V4 DSpark on AMD HIP platform: 本 PR 激活的 suite 正是 #30964 注册的；#30964 合并评审要求 MI35x + unified_kv_triton + GSM8K / DSpark 验收，本 PR 补上了这条执行链路。