

PR #25215 完整报告

sgl-project/sclang

[Docker] Fix several dependencies in Dockerfile

合并时间: 2026-05-14 11:18

原文链接: <http://prhub.com.cn/sgl-project/sclang/pull/25215>

执行摘要

- 一句话: 修复 Dockerfile 中 DeepEP 依赖和版本
- 推荐动作: 建议测试 CUDA 12 环境下使用新 DeepEP 仓库的镜像是否能正常构建和运行。

功能与动机

跟随 #25113 修复, 确保 Docker 构建镜像时能正确获取并构建 DeepEP 依赖, 避免因分支或版本错误导致构建失败。

实现拆解

1. 移除废弃变量: 删除 GRACE_BLACKWELL_DEEPEP_BRANCH 参数。
2. 依赖版本升级: 将 SGL_DEEP_GEMM_VERSION 从 0.0.1 改为 0.1.0。
3. 重构 DeepEP 克隆逻辑: 当 GRACE_BLACKWELL=1 时, 根据 CUDA_VERSION 主版本选择不同的仓库:
 - CUDA 12 使用 <https://github.com/fzyzcjy/DeepEP.git> 的 gb200_blog_part_2 分支。
 - 其他 CUDA 版本使用 <https://github.com/deepseek-ai/DeepEP.git> 的 hybrid-ep 分支, 并固定到指定 commit。
4. 超时宏修改: 保持原有对 NUM_CPU_TIMEOUT_SECS 和 NUM_TIMEOUT_CYCLES 的增强。

关键文件:

- docker/Dockerfile (模块 部署脚本; 类别 infra; 类型 infrastructure): 所有依赖版本与克隆逻辑的修改都集中在此文件

关键符号: 未识别

关键源码片段

docker/Dockerfile

所有依赖版本与克隆逻辑的修改都集中在此文件

```
# Dockerfile 变更片段
```

```
# 移除废弃的 GRACE_BLACKWELL_DEEPEP_BRANCH 参数  
ARG GRACE_BLACKWELL=0
```

```

ARG HOPPER_SBO=0
# -ARG GRACE_BLACKWELL_DEEPEP_BRANCH=hybrid-ep # 已移除
ARG HOPPER_SBO_DEEPEP_COMMIT=9f2fc4b3182a51044ae7ecb6610f7c9c3258c4d6
ARG DEEPEP_COMMIT=9af0e0d0e74f3577af1979c9b9e1ac2cad0104ee
ARG BUILD_AND_DOWNLOAD_PARALLEL=8
ARG SGL_KERNEL_VERSION=0.4.2.post1
ARG SGL_VERSION
# 升级 deep_gemm 版本
ARG SGL_DEEP_GEMM_VERSION=0.1.0 # 之前是 0.0.1

# .....

# 在 deepep_builder 阶段, 重构 Grace Blackwell 的 DeepEP 克隆逻辑
FROM torch_deps AS deepep_builder
ARG CUDA_VERSION
ARG BUILD_AND_DOWNLOAD_PARALLEL
ARG GRACE_BLACKWELL
# -ARG GRACE_BLACKWELL_DEEPEP_BRANCH # 已移除
ARG HOPPER_SBO
ARG HOPPER_SBO_DEEPEP_COMMIT
ARG DEEPEP_COMMIT

WORKDIR /build

# Clone DeepEP
RUN set -eux; \
    if [ "$GRACE_BLACKWELL" = "1" ]; then \
        # 根据 CUDA 主版本选择不同的仓库和分支
        if [ "${CUDA_VERSION%%.*}" = "12" ]; then \
            # CUDA 12 使用 fzyzcjy 的仓库, gb200_blog_part_2 分支
            git clone https://github.com/fzyzcjy/DeepEP.git && \
            cd DeepEP && \
            git checkout gb200_blog_part_2 && \
            # 调整超时宏 (保持原有修改)
            sed -i 's/#define NUM_CPU_TIMEOUT_SECS 100/#define NUM_CPU_TIMEOUT_SECS 1000/' csrc/kernels/configs.cuh && \
            sed -i 's/#define NUM_TIMEOUT_CYCLES 20000000000ull/#define NUM_TIMEOUT_CYCLES 20000000000ull/' csrc/kernels/configs.cuh && \
            cd .. ; \
        else \
            # 其他 CUDA 版本使用 deepseek-ai 的主仓库, hybrid-ep 分支, 固定 commit
            git clone https://github.com/deepseek-ai/DeepEP.git -b hybrid-ep && \
            cd DeepEP && \
            git checkout d28bd676c2120573c9f1425f0c16c39faa4117e6 && \
            sed -i 's/#define NUM_CPU_TIMEOUT_SECS 100/#define NUM_CPU_TIMEOUT_SECS 1000/' csrc/kernels/configs.cuh && \
            sed -i 's/#define NUM_TIMEOUT_CYCLES 20000000000ull/#define NUM_TIMEOUT_CYCLES 20000000000ull/' csrc/kernels/configs.cuh && \
            cd .. ; \
    fi

```

```
fi; \  
elif [ "$HOPPER_SBO" = "1" ]; then \  
  # 原有 Hopper SBO 逻辑不变  
  git clone https://github.com/deepseek-ai/DeepEP.git -b antgroup-opt && \  
  cd DeepEP && \  
  git checkout ... && \  
  cd .. ; \  
fi
```

评论区精华

无审核评论，仅提交者自行合并。

- 暂无高价值评论线程

风险与影响

- 风险：低风险。仅修改 Docker 构建过程，不影响运行时逻辑。但若新 DeepEP 仓库分支不稳定或与 CUDA 版本不兼容，可能导致镜像构建失败。
- 影响：影响 Docker 镜像构建，特别是针对 Grace Blackwell 架构的镜像。普通用户无影响。
- 风险标记：外部依赖变更

关联脉络

- PR #25113（推断）之前引入 DeepEP 相关分支的 PR: 本 PR 修复的工具链改动源自该 PR