

PR #49516 完整报告

vllm-project/vllm

[ROCm][CI] Keep native datasets cache off shared NFS

合并时间: 2026-07-27 11:31

原文链接: <http://prhub.com.cn/vllm-project/vllm/pull/49516>

执行摘要

- 一句话: AMD CI 原生测试数据集改用本地缓存路径
- 推荐动作: 可直接合并, 变更简单且经过审核。可作为 CI 基础设施缓存隔离的参考做法。

功能与动机

AMD CI 原生测试依赖 HuggingFace Datasets, 而共享 NFS 文件系统不支持 POSIX 锁, 导致测试随机失败。PR body 中引用了失败的 Buildkite 构建链接, 并说明需要将 Datasets 缓存放在 pod 本地存储。

实现拆解

1. 在 `.buildkite/scripts/hardware_ci/run-amd-test.sh` 脚本的 `initialize_native_environment` 函数中, 新增 `HF_DATASETS_CACHE` 环境变量指向 `${native_root}/cache/huggingface/datasets`, 即 pod 本地目录。
2. 将该变量加入 `export` 语句, 同时保留原有的 `HF_HOME` 共享路径不变, 实现 Hub 缓存持久化与 Datasets 缓存本地化分离。
3. 在 `mkdir -p` 命令中添加 `${HF_DATASETS_CACHE}` 目录的创建, 确保目录存在。
4. 修改对应的注释说明更改理由。

关键文件:

- `.buildkite/scripts/hardware_ci/run-amd-test.sh` (模块 部署脚本; 类别 `infra`; 类型 `infrastructure`): 唯一修改的文件, 新增 `HF_DATASETS_CACHE` 环境变量以分离数据集缓存路径, 解决 NFS 锁问题。

关键符号: `initialize_native_environment`

关键源码片段

`.buildkite/scripts/hardware_ci/run-amd-test.sh`

唯一修改的文件, 新增 `HF_DATASETS_CACHE` 环境变量以分离数据集缓存路径, 解决 NFS 锁问题。

```
# 初始化原生测试环境
initialize_native_environment() {
    # ... 其他缓存路径设置 ...
```

```

: "${HF_HOME:=/home/buildkite-agent/huggingface}" # 共享 NFS 上的 HF 主缓存, 持久化保存
Hub 模型
# datasets uses POSIX locks that are unsupported by the shared HF NFS cache.
# Keep processed datasets job-local while retaining the persistent Hub cache.
HF_DATASETS_CACHE="${native_root}/cache/huggingface/datasets" # 本地 Pod 存储, 避免
POSIX 锁冲突
# ...

# 导出环境变量 (新增 HF_DATASETS_CACHE)
export HF_HOME HF_DATASETS_CACHE HF_HUB_DOWNLOAD_TIMEOUT HF_HUB_ETAG_
TIMEOUT

# 确保所有缓存目录存在 (新增 ${HF_DATASETS_CACHE})
mkdir -p "${TMPDIR}" \
    "${TORCHINDUCTOR_CACHE_DIR}" \
    "${TRITON_CACHE_DIR}" \
    "${VLLM_CACHE_ROOT}" \
    "${XDG_CACHE_HOME}" \
    "${HF_HOME}" \
    "${HF_DATASETS_CACHE}" || return 1
}

```

评论区精华

审核者 [tjtanaa](#) 表示该方案是基础设施层面的解决, 无 spec 和 CI 环境方面的深入知识, 但接受了此变更, 并指出当前仅针对一种 GPU 集群类型, 后续可扩展到其他集群。

- 无 review 讨论线程 (other): 无。

风险与影响

- 风险: 风险极低。变更仅涉及 CI 脚本中的环境变量设置和目录创建, 不改变任何运行时逻辑。可能的风险是缓存目录路径在 pod 中不存在或空间不足, 但通过 `mkdir -p` 保证。
- 影响: 仅影响 AMD CI 原生测试环境。预期解决因 NFS POSIX 锁导致的测试失败, 提高 CI 稳定性。对其他平台、生产环境无影响。
- 风险标记: 变更范围极小, 仅 CI 脚本改动

关联脉络

- PR #49901 [CI/Build] Refresh tags before building macOS wheel: 同属 CI 基础设施改进, 但功能无关。