

PR #41032 完整报告

vllm-project/vllm

[Docker] Install numactl CLI in CUDA runtime image

合并时间: 2026-04-28 01:58

原文链接: <http://prhub.com.cn/vllm-project/vllm/pull/41032>

执行摘要

- 一句话: Docker 中安装 numactl 命令行工具
- 推荐动作: 简单的依赖修复 PR, 可直接合并。建议后续检查 dev 阶段是否需要同步添加。

功能与动机

用户在使用 vLLM 容器时, 若启用了 NUMA 绑定功能, 会报错 "numactl is required for NUMA binding but is not installed or not available on PATH"。PR body 和 commit message 明确指出: vllm-base image installs libnuma-dev but not the numactl binary, so NUMA binding fails at runtime。

实现拆解

1. 在 CUDA 运行时镜像安装阶段添加 numactl 包: 在 docker/Dockerfile 的 CUDA 运行时依赖安装命令中, 将原先单独一行的 libnuma-dev 拆为两行, 新增 numactl 包。两行之间添加了注释 # numactl CLI for NUMA binding at runtime 说明安装目的。

关键文件:

- docker/Dockerfile (模块 部署脚本; 类别 infra; 类型 infrastructure): 唯一变更文件, 在 CUDA 运行时镜像的 apt-get install 命令中添加 numactl 包。

关键符号: 未识别

关键源码片段

docker/Dockerfile

唯一变更文件, 在 CUDA 运行时镜像的 apt-get install 命令中添加 numactl 包。

```
# 在 CUDA 运行时依赖安装命令中, 将 libnuma-dev 拆为两行并新增 numactl
# 原: libnuma-dev && \
# 改为:
libnuma-dev \
# numactl CLI for NUMA binding at runtime
numactl && \
```

评论区精华

gemini-code-assist[bot] 建议也向 `dev` 阶段的镜像中添加 `numactl`，因为 `dev` 镜像是贡献者和用户的主要开发环境。该建议未被采纳或跟进（PR 已合并未在 `dev` 阶段添加）。

- 建议在 `dev` 阶段也添加 `numactl` (design): 未采纳，PR 仅修改运行时镜像阶段。

风险与影响

- 风险：低风险。仅在 Docker 镜像中新增一个包安装，不会影响现有功能或引入回归。但未在 `dev` 阶段同步安装可能导致开发环境仍缺少该二进制。
- 影响：影响范围：仅限使用 CUDA 运行时镜像的用户。这些用户现在可以在容器内正常使用 NUMA 绑定功能，无需手动安装 `numactl`。影响程度：低，仅涉及 Docker 构建环节。
- 风险标记：`dev` 阶段未同步安装 `numactl`

关联脉络

- PR #20384 fastsafetensors libnuma-dev issue reference: Docker 注释中引用了该 issue，指出 `libnuma-dev` 是为了修复 `fastsafetensors` 的问题。