

PR #47139 完整报告

vllm-project/vllm

[CI/Build] Bump PyNvVideoCodec version

合并时间: 2026-06-30 21:38

原文链接: <http://prhub.com.cn/vllm-project/vllm/pull/47139>

执行摘要

- 一句话: 升级 PyNvVideoCodec 依赖版本
- 推荐动作: 该 PR 是简单的依赖升级, 变更小, 风险中等。建议关注 pynvcodec 2.1.0 的后续修复版本, 或与上游协调发布 2.0.4 的 aarch64 轮子。若发现视频解码相关问题, 可考虑将依赖改为可选或增加版本范围。

功能与动机

PyNvVideoCodec 2.0.4 在 [manylinux_2_28_aarch64](#) 下缺少预编译 wheel, 导致 vLLM 在 ARM 架构上的 CI 或用户部署中无法安装该依赖。PR 关联到 PR#44465 的 issue 评论, 旨在通过升级到 2.1.0 获取更完善的架构支持。

实现拆解

关键文件:

- requirements/cuda.txt (模块 依赖管理; 类别 infra; 类型 dependency-upgrade) : 唯一变更文件, 将 PyNvVideoCodec 从 2.0.4 改为 2.1.0, 解决 aarch64 缺少 wheel 的问题。

关键符号: 未识别

关键源码片段

requirements/cuda.txt

唯一变更文件, 将 PyNvVideoCodec 从 2.0.4 改为 2.1.0, 解决 aarch64 缺少 wheel 的问题。

```
# requirements/cuda.txt ( 片段 )
torch==2.11.0
torchaudio==2.11.0
# These must be updated alongside torch
torchvision==0.26.0
# Required for phi3v processor.
# PyNvVideoCodec version bumped from 2.0.4 to 2.1.0 for aarch64 wheel support.
PyNvVideoCodec==2.1.0
# FlashInfer should be updated together with the Dockerfile
flashinfer-python==0.6.13
flashinfer-cubin==0.6.13
```

评论区精华

讨论中品牌方 [brandonpelfrey](#) 指出 2.1.0 版本存在已知问题，并询问是否可以为 2.0.4 补充 [aarch64](#) 的 wheel 以回退版本。[Harry-Chen](#) 表示如果能提供 2.0.4 的 [manylinux_2_28_aarch64](#) 轮子，则可以回退。同时有其他贡献者询问 2.1.0 为何缺少 cp313 及以上的 wheel，[Harry-Chen](#) 表示可能需要将依赖设为可选。最终该 PR 仍合并了 2.1.0 的升级。

- PyNvVideoCodec 2.1.0 的已知问题及 wheel 架构支持 (other): 由于尚未获得 2.0.4 的 [aarch64](#) 轮子，仍合并了 2.1.0 的升级。对于 cp313+ 的支持问题，考虑将依赖设为可选。

风险与影响

- 风险: PyNvVideoCodec 2.1.0 存在第三方报告的已知 issue，可能影响某些场景下的视频解码稳定性或功能。如果用户使用受影响的配置，视频处理可能出现错误。同时，2.1.0 缺乏 cp313+ 的 wheel，可能使 Python 3.13 用户无法安装该依赖，需依赖旧版本或降级 Python。
- 影响: 直接影响使用 NVIDIA GPU 和 PyNvVideoCodec 进行视频解码的用户，尤其是 ARM 架构 ([aarch64](#)) 用户现在可以安装该依赖。对使用 `--media-io-kwarg.video.backend=pynvvideocodec` 的用户功能正常 (PR 测试通过)。对于 Python 3.13+ 用户，依赖可能不可用。
- 风险标记: 依赖版本更新，上游已知问题，缺少 Python 3.13 支持

关联脉络

- PR #44465 原始涉及视频解码相关的 PR: 本 PR 的动机源自该 PR 中关于 PyNvVideoCodec 2.0.4 缺少 [aarch64](#) 轮子的评论。