PR #26879 完整报告

sgl-project/sglang

[AMD] Pin compressed-tensors==0.15.0 to fix ROCm nightly build

合并时间：2026-06-01 17:33

原文链接：http://prhub.com.cn/sgl-project/sglang/pull/26879

执行摘要

- 一句话：精确锁定 compressed-tensors 版本修复 ROCm 构建
- 推荐动作：值得立即合入以恢复 ROCm 每日构建。后续可关注 ROCm 基础镜像更新，适时解除压缩。

功能与动机

ROCm 每日构建自 2026-05-29 起失败，根因是 compressed-tensors 0.16.0 新增 torch>=2.10.0 约束，与 ROCm 基础镜像 torch 2.9.1 冲突。之前的开区间上限修复（#26591）导致 pip 回溯到古老的 setuptools 4.0.1 sdist，构建崩溃。

实现拆解

将 python/pyproject_other.toml 中 srt_hip 依赖组的 compressed-tensors 约束从开区间 <0.16.0 改为精确版本 ==0.15.0，并更新注释说明原因。动机：精确版本让 pip 解析器只有一个候选，避免回溯到不可构建的旧版 setuptools。已验证 ROCm 7.0 和 7.2 构建通过。

关键文件：

- python/pyproject_other.toml（模块 依赖配置；类别 config；类型 configuration）：唯一变更文件，修改 compressed-tensors 依赖版本约束，影响 ROCm 构建依赖解析。

关键符号：未识别

关键源码片段

python/pyproject_other.toml

唯一变更文件，修改 compressed-tensors 依赖版本约束，影响 ROCm 构建依赖解析。

```
# python/pyproject_other.toml
# ... 其他依赖项 ...

# HIP (Heterogeneous-computing Interface for Portability) for AMD
# => base docker rocm/vllm-dev:20250114, not from public vllm whl
srt_hip = [
  "sglang[runtime_common]",
  "torch",
  "petit_kernel==0.0.2",
```

```
    "wave-lang==3.8.2",
    # Pin to 0.15.0: 0.16.0 needs torch>=2.10 (incompatible with ROCm torch
    # 2.9.1). An open-ended `<0.16.0` made pip backtrack into an unbuildable
    # ancient setuptools sdist; an exact pin keeps the resolver converging.
    "compressed-tensors==0.15.0",
]
# ... 其他依赖组 ...
```

评论区精华

无 review 评论。

- 暂无高价值评论线程

风险与影响

- 风险：风险极低：仅修改一行依赖版本约束，且已在两个 ROCm 版本上验证通过。长期风险是当 ROCm 基础镜像升级到 torch>=2.10 后需解除此 pin，但 PR 已注明该限制。
- 影响：直接影响 AMD ROCm 平台的 Docker 构建流程，修复 nightly 构建失败。用户无感知，仅在安装 sglang[all] 或 srt_hip 时使用指定版本。
- 风险标记：依赖版本锁定需在 ROCm 基础镜像升级后解除

关联脉络

- PR #26591 hotfix: pin compressed-tensors<0.16.0 for ROCm: 前一次修复，因使用开区间上限导致解析器回溯，本 PR 在其基础上改为精确版本。