

PR #1700 完整报告

THUDM/slime

fix: auto-detect GPUs in qwen3-4b script

合并时间: 2026-03-22 16:27

原文链接: <http://prhub.com.cn/THUDM/slime/pull/1700>

执行摘要

修复 qwen3-4B 训练脚本中 GPU 数量硬编码问题，通过自动检测 nvidia-smi 输出实现动态配置，简化用户操作并避免配置错误，属于重要的 bugfix 改进。

功能与动机

动机源于 Issue #148，用户报告在 4xH100 机器上运行时因脚本中 GPU 数量未从 8 更新为 4 导致失败，错误日志不明确。本 PR 旨在自动化 GPU 数量检测，减少手动配置需求，引用 PR body: "derive NUM_GPUS from nvidia-smi when available - reuse NUM_GPUS for ray start and actor GPU flags"。

实现拆解

改动集中于 `scripts/run-qwen3-4B.sh` 文件：

- 添加 GPU 检测逻辑：使用 `nvidia-smi -L` 命令获取 GPU 数量，如果命令不可用则设为 0。
- 设置 `NUM_GPUS` 变量：优先使用用户定义值（通过环境变量），否则使用检测值，若仍为空或小于等于 0 则默认设为 8。
- 替换硬编码参数：将 `ray start --num-gpus 8` 和 `--actor-num-gpus-per-node 8` 中的 8 替换为 `${NUM_GPUS}`。

关键代码块示例：

```
if command -v nvidia-smi >/dev/null 2>&1; then
  DETECTED_GPUS=$(nvidia-smi -L 2>/dev/null | wc -l | tr -d ' ')
else
  DETECTED_GPUS=0
fi
NUM_GPUS=${NUM_GPUS:-${DETECTED_GPUS}}
if [ -z "$NUM_GPUS" ] || [ "$NUM_GPUS" -le 0 ]; then
  NUM_GPUS=8
fi
```

评论区精华

无实质性 review 讨论。

风险与影响

- 风险：依赖 `nvidia-smi` 命令，若不可用或输出格式变化可能导致检测错误；默认值 8 假设可能不适用于所有 GPU 环境，引发配置不匹配。
- 影响：用户无需手动修改脚本，自动化提升易用性；仅影响 qwen3-4B 模型的训练流程，减少配置错误导致的运行失败。

关联脉络

与近期 PR #1719（修复 Qwen3-235B-A22B 脚本）和 #1689（修复 shell 脚本变量引用）类似，都属于 shell 脚本的 bugfix，反映团队在持续优化模型启动脚本的健壮性和配置自动化。