

PR #6942 完整报告

verl-project/verl

[hardware, perf] fix: update AMD GPU FLOPS values

合并时间: 2026-07-06 20:47

原文链接: <http://prhub.com.cn/verl-project/verl/pull/6942>

执行摘要

- 一句话: 更新 AMD GPU FLOPS 数据表格
- 推荐动作: 该 PR 为一次简单的硬件参数修正, 改动小且正确性明确, 值得快速合并。开发者若涉及 AMD GPU 的性能评测或 MFU 监控, 应关注此变更以确保数据准确性。

功能与动机

AMD 官方 BF16/FP16 稠密峰值在 MI300X 上公布约为 1307 TFLOPS, 此前使用的 1336e12 有偏差; MI350X 和 MI355X 未被收录, 导致在这些 GPU 上运行训练时 MFU 无法正确计算 (`get_device_flops()` 返回 `inf`, MFU 显示为 0)。详见 PR body。

实现拆解

仅修改 `verl/utils/flops_counter.py` 中的 `_DEVICE_FLOPS` 字典:

1. 将 "MI300X": 1336e12 修正为 "MI300X": 1307e12。
2. 新增 "MI350X": 2300e12。
3. 新增 "MI355X": 2500e12。无其他文件变更, 未涉及配置、测试或文档更新。

关键文件:

- `verl/utils/flops_counter.py` (模块 工具模块; 类别 source; 类型 core-logic): 唯一变更文件, 修正了 MI300X 峰值 FLOPS 并新增 MI350X/MI355X 条目, 直接影响 AMD GPU 上 MFU 计算的正确性。

关键符号: 未识别

关键源码片段

`verl/utils/flops_counter.py`

唯一变更文件, 修正了 MI300X 峰值 FLOPS 并新增 MI350X/MI355X 条目, 直接影响 AMD GPU 上 MFU 计算的正确性。

```
# verl/utils/flops_counter.py
```

```
_DEVICE_FLOPS = {  
    "CPU": 448e9,  
    "GB200": 2.5e15,
```

```
"B200": 2.25e15,  
# MI300X: 修正为 AMD 官方 BF16/FP16 稠密峰值 ~1307 TFLOPS  
"MI300X": 1307e12,  
# 新增 MI350X 和 MI355X ( 官方 BF16 稠密峰值 )  
"MI350X": 2300e12, # 2.3 PFLOPS  
"MI355X": 2500e12, # 2.5 PFLOPS  
"H100": 989e12,  
"H800": 989e12,  
"H200": 989e12,  
"A100": 312e12,  
"A800": 312e12,  
"L40S": 362.05e12,  
"L40": 181.05e12,  
"A40": 149.7e12,  
"L20": 119.5e12,  
"H20": 148e12,  
"910B": 354e12,  
"A2G3": 354e12,  
"Ascend950DT": 432e12,  
"Ascend910": 354e12,  
"RTX 3070 Ti": 21.75e12,  
}
```

评论区精华

无 review 讨论。PR 被 wuxibin89 直接批准，仅有一条来自 [gemini-code-assist\[bot\]](#) 的自动评论，无实际反馈。

- 暂无高价值评论线程

风险与影响

- 风险：风险极低。仅修改静态 FLOPS 常量字典，修正值和新增条目均来自 AMD 官方公布数据。可能存在的风险：若官方数据后续更新或不同精度下峰值不同，需同步更新；但当前变更已注明使用 BF16/FP16 稠密峰值，标注清楚即可。
- 影响：影响范围小，仅影响 AMD MI300X、MI350X、MI355X GPU 上的 MFU 计算。修正后 MI300X 用户会看到 MFU 从 0（或错误值）变为更准确的值；新增的 MI350X 和 MI355X 用户可正常计算 MFU。对 NVIDIA、Ascend 等其他硬件无影响。
- 风险标记：暂无

关联脉络

- 暂无明显关联 PR