

PR #54549 完整报告

vllm-project/vllm

[CI] Mark 1-GPU L4 test steps with device: I4 for EKS migration

合并时间: 2026-09-01 05:56

原文链接: <http://prhub.com.cn/vllm-project/vllm/pull/54549>

执行摘要

- 一句话: 8 个 1-GPU L4 测试步骤标记 device: I4 迁移到 EKS
- 推荐动作: 值得快速浏览 (约 5 分钟) 以理解 vLLM CI 的 EKS 迁移模式, 但无需精读源码。关注两个设计点: 一是 device: I4 + num_devices: 1 作为路由契约的最小侵入式标记方式——只需在 YAML 里加键即可让生成器完成队列路由与资源缩放; 二是迁移边界的取舍逻辑 (torch-abi 因审计 CUDA wheel 需求被排除、2-node 因多机特性留在 ASG), 这类 "什么不能迁" 的决策比 "迁什么" 更有参考价值。若后续你所在团队也在做 CI 容器化迁移, 这个 PR 连同 #54326、ci-infra#499 可作为一个完整的渐进迁移案例。

功能与动机

PR body 明确说明这是 #54326 的后续推进: "Follow-up to #54326 (which moved the 2/4-GPU L4 steps to the EKS I4-k8s queue): mark the remaining 1-GPU L4 steps device: I4 so they migrate too." 动机是 CI 基础设施从 ASG 全面迁移到 EKS, 以提升 GPU 资源打包效率——"On EKS these pack 4-per-node (MostAllocated): the whole 1-GPU class occupies ~2-3 g6.12xlarge at typical concurrency vs 208 single-GPU ASG slots today"。作者还特别声明这是迁移设计验证后的收尾工作, 2/4-GPU 类别已在 I4-ci 集群上稳定运行两天 (45/45 与 31/31 验证构建通过)。

实现拆解

1. 建立路由契约

ci-infra#499 的 pipeline 生成器 (已合并) 规定: 任何 Buildkite 步骤只要声明 device: I4, 就会被路由到 EKS I4-k8s 队列, 并按 num_devices 缩放请求的 GPU 资源。因此本 PR 无需改动生成器, 只需在测试步骤的 YAML 定义中补上这两个键。

2. 标记 8 个 1-GPU 步骤 (涉及 5 个文件)

- .buildkite/test_areas/kernels.yaml: 给 Attention Kernels Shard、Quantization Kernels Shard、MoE Kernels Shard 三个步骤各追加 device: I4 + num_devices: 1 (每个 shard 独占 1 GPU, parallelism 不变)。
- .buildkite/test_areas/engine.yaml: 给 engine-1-gpu (V1 Engine) 步骤追加 device: I4 + num_devices: 1。

- `.buildkite/test_areas/models_language.yaml`: 给 Granite Language Model Compatibility 步骤追加 `device: l4 + num_devices: 1`。
- `.buildkite/test_areas/pytorch.yaml`: 给 PyTorch Compilation Passes 与 PyTorch Fullgraph CUDAGraph Compatibility 两个步骤各追加 `device: l4 + num_devices: 1`。
- `.buildkite/test_areas/distributed.yaml`: 给 Sharded RDT Weight Transfer 步骤追加 `device: l4` (该步骤原本已有 `num_devices: 1`)。

3. 边界决策：有意排除两个步骤

- Torch Stable ABI Audit 不标记：它被 #54468 恢复到 `gpu_1_queue`，因为该步骤必须审计 CUDA wheel，不能迁移到 EKS 环境。
- 2-node 步骤不标记：multi-node 测试继续留在 ASG，EKS 迁移范围当前只覆盖单机内的单 GPU/ 多 GPU 场景。

4. 验证配套（无代码测试，属 CI 配置自检）

作者声称做了两类验证：对全部 `test_areas` 文件做 YAML 解析检查；用 `ci-infra main` 生成完整 pipeline (`RUN_ALL=1`) 核对路由结果——39 个步骤进入 `l4-k8s`，`torch-abi` 仍进 `gpu_1_queue`，2-node 进 `gpu_4_queue`，其余步骤无变化。此外 2/4-GPU 同规格 pod 已在 `l4-ci` 集群上 soak 两天作为运行时证据。

关键文件：

- `.buildkite/test_areas/kernels.yaml` (模块 CI 配置；类别 config；类型 configuration)：改动量最大：Attention / Quantization / MoE 三个内核测试分片步骤各新增 `device: l4 + num_devices: 1`，是本次迁移中测试覆盖最核心的文件。
- `.buildkite/test_areas/engine.yaml` (模块 CI 配置；类别 config；类型 configuration)：V1 Engine 步骤 (key: `engine-1-gpu`) 是 vLLM v1 引擎核心测试，迁移后运行在 EKS 上。
- `.buildkite/test_areas/models_language.yaml` (模块 CI 配置；类别 config；类型 configuration)：Granite Language Model Compatibility 步骤因对 Triton SSD 自动调优敏感而特意保留在 L4 上，迁移时需确保 EKS 环境不改变其行为。
- `.buildkite/test_areas/pytorch.yaml` (模块 CI 配置；类别 config；类型 configuration)：PyTorch Compilation Passes 与 Fullgraph CUDAGraph Compatibility 两个编译相关步骤迁移，涉及 `torch.compile` 与 `cudagraph` 行为，对环境差异敏感。
- `.buildkite/test_areas/distributed.yaml` (模块 CI 配置；类别 config；类型 configuration)：Sharded RDT Weight Transfer 步骤原本已有 `num_devices: 1`，本次仅补 `device: l4` 即完成迁移标记。

关键符号：未识别

评论区精华

该 PR 没有任何实质性的技术 review 讨论：`review_comments_count` 为 0，唯一的系统评论来自 `claude[bot]`，内容是自动提示本仓库配置为人工 review 模式，无技术价值。因此决策信息全部来自 PR body 中作者的自述验证：生成器路由核对结果、2/4-GPU 类别的两天 soak 数据，以及排除 `torch-abi` 与 2-node 步骤的明确理由。

- 暂无高价值评论线程

风险与影响

- 风险:

1. 同节点邻居干扰: EKS 侧使用 MostAllocated 策略将 4 个作业打包到同一个 g6.12xlarge 节点, 邻居作业的 CPU/ 内存 /GPU 争用可能让原本稳定的测试 (如 Granite 兼容性、PyTorch 编译 passes) 出现新的 flaky, 而这些步骤覆盖 kernels、engine、compile 等核心测试区域。
2. 环境差异: EKS 的 I4 pod 规格与 ASG 物理机可能存在驱动、容器运行时、内核参数的细微差异; 虽然 2/4-GPU 类别已 soak 两天, 但 1-GPU 类别包含 8 个不同测试套件, 覆盖行为不同。
3. 配置契约耦合: .buildkite/test_areas 下的步骤依赖 ci-infra 生成器对 device: I4 键的解释, 生成器或队列命名变化会静默影响这些步骤的路由, 且本 PR 无自动化测试守护这一契约。
4. 缺少人工 review: 全部改动由作者单方提交合并, 虽然改动机械 (15 行纯新增), 但影响的是 CI 主干稳定性。- 影响: 影响范围限于 CI 基础设施, 不涉及 vLLM 运行时、用户 API 或推理性能。对开发者而言, 这 8 个测试步骤的触发条件、测试内容、超时配置均未改变, 唯一变化是物理执行位置从 ASG 迁移到 EKS, 理论上行为应完全一致。对 CI 系统而言, 资源效率提升显著: 整个 1-GPU 类别从 208 个 ASG 单 GPU 槽位缩减为约 2-3 个 g6.12xlarge 节点 (典型并发下), 配合此前 #54326 的 2/4-GPU 迁移, L4 测试负载已基本完成 EKS 化。对团队而言, 这是 vLLM CI 演进序列中的一环, 后续其他 GPU 类别大概率会沿用同一 device: <型号> 标记模式。- 风险标记: CI 基础设施迁移, 调度行为变化, 缺少人工 review, 配置契约耦合, 核心测试区域

关联脉络

- PR #54326 [CI] Move 2/4-GPU L4 test steps to EKS I4-k8s queue: 本 PR 的直接前序: 已把 2/4-GPU L4 步骤迁移到 EKS I4-k8s 队列并完成两天 soak, 本 PR 将同一模式推广到 1-GPU 类别。
- PR #54468 [CI] Restore Torch Stable ABI Audit to gpu_1_queue: torch-abi 步骤被 #54468 恢复回 gpu_1_queue, 本 PR 因此刻意不给它加 device: I4, 两者共同界定了 EKS 迁移的边界。
- PR #54365 [CI] Exclude kv_transfer changes from broad spec-decode/kernels/multimodal triggers: 同属 .buildkite/test_areas 配置演进线, 展示 vLLM CI 正在系统性地调整测试区域 YAML 的触发与路由规则。