# PR #2356 完整报告

radixark/miles

Replace all the `.sh` launch scripts with `.py` launch script

合并时间：2026-08-12 12:51

原文链接：http://prhub.com.cn/radixark/miles/pull/2356

## 执行摘要

- 一句话：用 Python 启动器全面替换 scripts/ 下全部 shell 启动脚本
- 推荐动作：值得精读，推荐顺序：先读 PR body 和 command_utils.py 的 *execute_train* / ssh_start_ray_workers，再看 run_qwen3_dense.py 的 _Recipe 合并模式与 run_nemotron_3_super_120b_a12b.py 的角色拆分。最大的方法论价值是用"记录 - 重放快照对比"验证大规模重写语义等价，这种验证方式可复用于其他 bash→python 迁移项目。

## 功能与动机

PR body 明确指出这是 #1837 中启动器改写任务的一环：模型配置早已迁移到 python（#1910），本 PR 完成启动器本身的迁移。同时清理了三个 shell 遗留怪癖（rl_data 目录不存在、HF_HOME 注入无效、ray start 参数不一致），并借机统一了目录与 GPU 参数约定。

## 实现拆解

1. 前置清理与合并基线：分支基于 yueming/script-bugfix，带入了 #2354（删除废弃模型启动器）与 #2355（修复快照冻结的 bug）的改动；再往 main 合入时解决了 DeepSeek V3 文档页、glm4-5 页面等多处冲突。
2. 用 _Recipe 表合并近重复配方：每个变体的差异化参数收敛为 _Recipe dataclass，ScriptArgs.recipe 属性按 --model-name / --topology 选择。例如 run_qwen3_dense.py 覆盖 6 个 Qwen3/Qwen3.5/Qwen3.6 dense 配方，run_kimi_k2.py 覆盖 Instruct/Thinking，run_qwen3_next_80b_a3b.py 覆盖 4node/single-node 两种拓扑。这消除了逐文件拷贝导致的参数漂移。
3. 基础设施收敛到 *execute_train*：所有启动器统一通过 U.*execute_train* 提交，前导逻辑（环境变量、ray 生命周期、PYTHONPATH、进程清理）由 command_utils 统一管理；新增 U.ssh_start_ray_workers 让四个多节点配方通过 before_ray_job_submit 共享 ssh 扇出逻辑。有意的差异包括：--num-gpus-per-node 总是显式传入，--colocate 时去掉 --rollout-num-gpus，wandb 统一走 get_default_wandb_args。
4. 修正两个 shell 遗留怪癖：GLM-4.5 与 Kimi-K2-Instruct 原本读取不存在的 <data-dir>/rl_data/，现统一为 hf_download_dataset 实际写入的 <data-dir>/aime-2024；gpt-oss 停止注入对 ray workers 无效的 HF_HOME=/workspace/hf_cache。
5. 特殊脚本的处理：run_glm45_355b_a32b_8node.py 作为新文件而非合并进 run_glm45_355b_a32b.py（两者是同一模型的不同实验，记录 diff 有 77 处差异）；

run_nemotron_3_super_120b_a12b.py 用 typer 双命令 train / worker 表达头 /worker 角色拆分；Kimi-K2 配方用 MILES_SCRIPT_EXTERNAL_RAY=1 表达外部已起 ray；run_qwen3_4b_npu.sh 因指向错误的 CUDA 启动器且无有效读者而被直接删除。

6. 测试与文档配套：tests/fast/launch_scripts/test_sh_harness.py 改指向幸存的 examples 启动器，移除 6 个 scripts/ 条目并调整发现下限；harness 的 CLEARED_ENV 补充冻结 MLP_SOCKET_IFNAME 与 MLP_WORKER_0_HOST；约 21 个文档页面改用 python 调用，新增 docs/user-guide/launch-script.md 替代旧的 bash 中心 walkthrough。

关键文件：

- scripts/run_qwen3_dense.py（模块 启动器；类别 source；类型 core-logic；符号 _Recipe, ScriptArgs, recipe, execute）：最具代表性的配方合并启动器：6 个 Qwen3/Qwen3.5/Qwen3.6 dense shell 脚本收敛为一个 _Recipe 表，按 --model-name 切换，覆盖了本 PR 的核心设计模式。
- scripts/run_glm45_355b_a32b_8node.py（模块 启动器；类别 source；类型 core-logic；符号 ScriptArgs, _cluster_env_vars, execute, main）：8 节点 GSPO 实验的独立启动器，与现有 Blackwell/GRPO 启动器差异 77 处故不合并；完整承载 NCCL/IB 20+ 环境变量注入与 hostfile ssh 扇出，是本次判断力最强的文件。
- scripts/run_nemotron_3_super_120b_a12b.py（模块 启动器；类别 source；类型 core-logic；符号 ScriptArgs, _wait_for_head_port, _wait_for_ray_gpus, _execute_train）：首次用 typer 双命令表达 head/worker 角色拆分，包含端口轮询与 GPU 就绪等待逻辑，是本次新增启动器中最复杂的控制流。
- scripts/run_qwen3_next_80b_a3b.py（模块 启动器；类别 source；类型 core-logic；符号 _Recipe, ScriptArgs, recipe, execute）：按 --topology 用 _Recipe 表区分 4node 生产布局与 single-node 演示布局，展示 colocate 与独立 rollout GPU 两种拓扑的切换。
- scripts/run_kimi_k2.py（模块 启动器；类别 source；类型 core-logic；符号 _Recipe, ScriptArgs, recipe, execute）：合并 Kimi-K2 Instruct 与 Thinking 两个配方，并引入 MILES_SCRIPT_EXTERNAL_RAY=1 表达外部已有 ray 集群的启动模式。
- scripts/run_qwen3_sft.py（模块 启动器；类别 source；类型 core-logic；符号 _Recipe, ScriptArgs, recipe, execute）：纯 SFT 配方（--debug-train-only）与 RL 配方共享同一套 execute_train 基础设施，并演示多节点 ssh 接入的 recipe gate。
- scripts/amd/run_qwen3_4b.py（模块 启动器；类别 source；类型 core-logic；符号 ScriptArgs, post_init, execute, main）：AMD 平台专用启动器，处理 HIP/CUDA 可见设备镜像、RAY_EXPERIMENTAL_NOSET_* 与按硬件型号推导 GPU 数，是平台差异的集中体现。
- miles/utils/external_utils/command_utils.py（模块 命令工具；类别 source；类型 core-logic；符号 ssh_start_ray_workers）：本次改动的公共设施：新增 ssh_start_ray_workers，把四个启动器反复手写的 ssh 扇出循环收敛为共享助手，是后续启动器复用的基础。
- tests/fast/launch_scripts/py_harness.py（模块 测试工具；类别 test；类型 test-coverage）：快照 harness 是等价性验证的基石；guapisolo 指出的 MLP_WORKER_0_HOST 冻结问题在这里修复，直接影响快照测试的环境隔离。

- docs/user-guide/launch-script.md（模块 文档；类别 docs；类型 documentation；符号 _Recipe, ScriptArgs, model_args）：以 python 启动器为中心的新启动脚本指南，替代旧的 bash 中心 walkthrough，是用户侧迁移的主要入口。
- docs/user-guide/training-script-walkthrough.md（模块 文档；类别 docs；类型 deletion；符号 check_reward_nonzero_std, pop_first）：旧的 bash 中心演练页被移除，符号化地宣告 shell 启动器时代的结束；其内容由 launch-script.md 承接。
- scripts/run-qwen3-next-80B-A3B.sh（模块 启动器；类别 other；类型 deletion）：被 run_qwen3_next_80b_a3b.py 取代的代表性 shell 启动器之一，删除它体现了本 PR 的核心动作。

关键符号：*execute*, main, ScriptArgs, _Recipe, recipe, ssh_start_ray_workers, _wait_for_head_port, _wait_for_ray_gpus, _execute_train, _cluster_env_vars

## 评论区精华

1. guapisolo 的快照隔离问题（已修复）：在 tests/fast/launch_scripts/py_harness.py 上指出 freeze_environment() 只清了 MLP_SOCKET_IFNAME，而四个新启动器也会读 MLP_WORKER_0_HOST，不冻结会导致快照测试随开发机环境漂移；提交 0fa40f7 已把 MLP_WORKER_0_HOST 加入 CLEARED_ENV。
2. Zhichenzzz 对数据集目录统一的建议：在 docs/models/glm/glm4-5.md 评论“we should unify the /root/datasets/rl_data and /root/datasets/”；提交 d584106 统一为 aime-2024 目录。
3. Zhichenzzz 对 ssh 扇出开关的疑问：在 run_qwen3_next_80b_a3b.py 问“only in this script using this gate”；结论文档说明 join_ray_workers 的额外 gate 用于排除单节点配方，后续提交又统一了四个启动器的开关命名。
4. Zhichenzzz 对 gpt-oss 后端的确认：在 run_gpt_oss_20b.py 问“the unique one?”，确认 bshd + fused 是该模型 sink attention 在 TE 中唯一可行的组合。

- 快照冻结缺少 MLP_WORKER_0_HOST 导致环境依赖 (testing): 提交 0fa40f7 将 MLP_WORKER_0_HOST 加入 CLEARED_ENV，并确认 178 个快照全部稳定通过。
- rl_data 与数据集目录统一 (design): 提交 d584106 统一为 hf_download_dataset 实际写入的 aime-2024 目录，GLM-4.5 与 Kimi-K2-Instruct 配方及文档全部修正。
- join_ray_workers 开关是否只在单一脚本中使用 (question): 作者澄清：带单节点配方的启动器需要该 gate 防止 ssh 扇出，四个多节点启动器最终统一命名为 join_ray_workers。
- gpt-oss 的 bshd + fused 后端是否唯一 (question): 确认该模型在 TE 中只能用 qkv-format bshd + fused 后端，且因此必须用静态 micro-batch。

## 风险与影响

- 风险：
  1. 删除面大且无法回滚：24 个 shell 启动器被删除，外部书签、内部笔记、旧文档链接可能失效；文档已同步更新但无法覆盖所有使用方。
  2. 快照等价性只覆盖 argv 与 runtime env：验证能防 flag 漂移，但不覆盖真实训练行为；run_glm45_355b_a32b_8node.py 的 20+ NCCL/IB 环境变量若作用于错误进程会影

响整机性能，run_nemotron_3_super_120b_a12b.py 的 _wait_for_ray_gpus 在集群未就绪时会等待 10 分钟再提交。

3. NPU 支持空缺：run_qwen3_4b_npu.sh 被删除且无替代，scripts/run_qwen3_4b_npu.py 自身有 /root/model 拼写错误、prepare() 结果未使用等缺陷，需 NPU 硬件验证；NPU 用户按旧路径会直接失败。
4. 目录与参数默认值变化：--model-dir / --data-dir / --output-dir 默认值改为 /root/models、/root/datasets、/root/shared_data，旧 ${BASE_DIR} 等路径不再内建；依赖集群注入的 msc miles 场景需通过 MILES_SCRIPT_* 覆盖。
5. ray dashboard 访问变化：ray start 不再传 --dashboard-host/--dashboard-port，远程 dashboard 访问方式改变（提交仍走默认 8265）。 - 影响：对用户而言，所有模型启动入口从 bash 变量风格切换为 typer flags，6 个 Qwen3 dense shell 脚本合并为一个入口；文档和启动方式都需要重新学习。对团队而言，净删除约 11000 行，启动器规则文件（#2357）也开始约束 python 启动器的书写风格，长期降低维护成本。影响范围覆盖 scripts/、miles/utils/external_utils/command_utils.py、tests/fast 快照与 harness、docs 约 21 页，共 115 个文件，是本仓库启动器体系的系统性重构。 - 风险标记：115 文件大改动，删除 24 个 shell 启动器，快照验证不覆盖真实训练行为，NPU 启动器空缺，路径默认值变更，ray dashboard 访问变化

## 关联脉络

- PR #2355 [fix] fix the bugs/outdated commands in .sh scripts and the corresponding snapshots: 同一 scripts/ 清理栈的中间 PR，修复 shell 启动器 bug 并重录快照；本 PR 以它的分支 yueming/script-bugfix 为 base。
- PR #2354 Delete the glm4-9B, mimo-7B, moonlight-16B and deepseek-r1 launch scripts: 同一清理栈的首个 PR，删除废弃模型启动器，为 .sh → .py 替换扫清障碍。
- PR #1910 Replace the model config shell scripts with python: 同属 #1837 的先行工作，把模型配置 shell 脚本替换为 python；本 PR 完成启动器本身，任务闭环。
- PR #2391 docs: replace DeepSeek V3/R1 page with a DeepSeek-V3.2 recipe: 本 PR 分支在合入 main 时发现 run_deepseek.py 仍需要 V3 页面，因此恢复了被 #2391 替换掉的页面，两者的文档决策相互影响。
- PR #2357 Add a rule file for the python launch and model scripts: 作为提交 98f568d 合入本分支，为新的 python 启动器 / 模型脚本补上书写规则，是本次重构的配套规范。