

# PR #7085 完整报告

PaddlePaddle/FastDeploy

[CI] Optimize test execution with single-GPU parallelism

合并时间: 2026-04-01 14:18

原文链接: <http://prhub.com.cn/PaddlePaddle/FastDeploy/pull/7085>

## 执行摘要

本次 PR 针对 FastDeploy 仓库的 CI 测试流程进行优化，通过引入单 GPU 并行执行和日志隔离机制，旨在提升 GPU 利用率和缩短 CI 周转时间，同时改善测试失败时的调试效率。变更覆盖 CI 工作流、测试执行脚本和多个 e2e 测试文件，影响中等，值得关注测试基础设施的工程师参考。

## 功能与动机

根据 PR 作者描述，当前单元测试大多顺序执行，导致 "GPU 利用率低和 CI 周转时间长"，且测试失败时 "日志分散，难以定位，增加调试成本"。因此，本次优化的核心目标是解决 CI 效率低下和日志管理混乱的问题，提升开发体验。

## 实现拆解

实现分为三个主要部分：

- CI 工作流集成：在 `.github/workflows/_unit_test_coverage.yml` 中新增日志上传步骤，当测试失败时自动打包日志并上传至存储，便于后续访问。
- 测试执行逻辑：在 `scripts/coverage_run.sh` 中添加 `classify_tests` 函数，基于规则（如文件路径、内容关键词）将测试分类为多 GPU（需顺序执行）和单 GPU（可并行）。同时，通过 `run_test_with_logging` 函数实现单 GPU 测试的并行分片执行和多 GPU 测试的顺序运行，并为每个测试创建独立日志目录（基于 `FD_LOG_DIR` 环境变量）。
- 测试文件适配：修改多个 e2e 测试文件（例如 `tests/e2e/test_EB_Lite_serving.py`），将硬编码日志路径替换为使用 `FD_LOG_DIR` 环境变量，确保日志隔离和一致性。

## 评论区精华

本 PR 未产生实质性 review 讨论，仅有两名 reviewer（ZhangYulongg 和 DDDivano）批准，无技术争议或设计权衡记录。

## 风险与影响

风险：

- 并行执行可能引发竞争条件，特别是对于端口或 GPU 内存敏感的测试。
- 分类逻辑依赖正则和 `grep` 规则，存在误判风险，可能导致资源冲突或测试失败。
- 日志路径变更可能破坏现有测试的日志依赖，需确保环境变量正确设置。

- 核心脚本 `scripts/coverage_run.sh` 改动较大，增加 bug 引入概率。

影响：

- 正面：CI 测试时间缩短，GPU 利用率提升，调试效率改善。
- 负面：需监控并行执行稳定性和日志兼容性，团队需适应新流程。

## 关联脉络

从近期历史 PR 看，相关 PR 包括：

- PR #7098（日志参数优化）：共享 Logging 标签，聚焦日志输出改进，与本 PR 的日志收集形成协同。
- PR #7104（CI 跳过逻辑调整）：同属 CI 领域，优化 CI 触发机制，与本 PR 的测试执行优化互补。
- PR #6731（单元测试补充）：涉及测试覆盖和 CI 测试，反映仓库对测试基础设施的持续投入。整体趋势表明，仓库正通过多项优化提升测试效率和可维护性。