

PR #21202 完整报告

sgl-project/sglang

docs: improve CI and testing documentation

合并时间: 2026-03-24 01:48

原文链接: <http://prhub.com.cn/sgl-project/sglang/pull/21202>

执行摘要

本 PR 主要改进 SGLang 项目的 CI 和测试文档，通过大幅更新 `test/README.md` 提供清晰指南，并辅以 CI workflow 调整、测试文件修复和代码清理，旨在提升开发者体验和 CI 稳定性，属于常规维护性变更。

功能与动机

PR 动机源于改善 CI 和测试文档的清晰度，如 PR body 所述: 'improve documentation for CI and tests'，特别是在 `test/README.md` 中提供更明确的指导，帮助开发者理解测试流程和 CI 流水线。此外，伴随的 CI 调整和测试修复旨在解决实际运行问题，如 CPU 测试失败和脚本路径不一致。

实现拆解

实现分为三大模块:

1. 文档更新: `test/README.md` 被重写，添加三阶段 CI 流水线示意图和详细表格，例如:
markdown - Stage A: 预检测测试 (~3分钟) - Stage B: 基础测试 (~30分钟) - Stage C: 高级测试 (~30分钟)
2. CI 和脚本调整: 将 AMD CI 脚本 (如 `ensure_vram_clear.sh`) 移动到 `scripts/ci/amd/` 目录，更新 `.github/workflows/` 文件中的路径引用，确保一致性。
3. 测试文件修复: 在 87 个测试文件中 (如 `python/sglang/jit_kernel/tests/test_*.py`) 添加 `import sys` 并用 `sys.exit(pytest.main(...))` 包裹调用，修复退出码问题；同时使用 `register_fake_if_exists` 改进 `mock` 机制，避免 CPU 测试失败。
4. 代码清理: 移除未使用的文件，如 `test/srt/experiment_runner.py` 和 `test/srt/compare.py`，简化代码库。

评论区精华

review 中仅有一次核心讨论:

```
gemini-code-assist[bot] 指出: 'pytest.main() returns an exit code but does not exit the process. To ensure that CI correctly captures test failures, you should wrap the call in sys.exit().'
```

作者 merrymercy 回复确认已在提交中修复，体现了对测试正确性的重视，无未解决疑虑。

风险与影响

- 风险：CI 脚本路径变更可能导致 workflow 引用错误（需验证 `.github/workflows/` 文件）；测试文件修改可能引入语法错误（已通过提交修复）；移除未使用代码可能影响遗留依赖（但无已知问题）。
- 影响：开发者获得更清晰的文档，降低学习成本；CI 系统通过修复测试执行问题提升可靠性；团队效率因标准化和清理而提高。影响范围覆盖所有贡献者和 CI 管道，程度为中等。

关联脉络

本 PR 是近期 CI 和测试改进趋势的一部分：

- 与 PR #21187 'ci: unify PR test suite naming' 直接相关，共同推动 CI 标准化。
- 与 PR #21200 '[NPU] bugfix for import sgl-kernel error' 共享代码修复（`hisparse_memory_pool.py`）。
- 同系列测试 PR（如 #21010、#21002）为本 PR 的文档更新提供上下文，显示项目正加强测试基础设施。整体揭示 SGLang 在提升开发体验和代码质量方面的持续演进。