# PR #33044 完整报告

sgl-project/sglang

[CI] Pin NCCL ports for GB300 PR tests

合并时间：2026-07-31 17:28

原文链接：http://prhub.com.cn/sgl-project/sglang/pull/33044

## 执行摘要

- 一句话：固定 GB300 PR 测试 NCCL 端口，消除 EADDRINUSE 竞争失败
- 推荐动作：值得快速浏览，尤其是对 CI 测试稳定性维护感兴趣的工程师。其设计决策——从共享默认端口常量推导基准值并分段偏移——比硬编码端口更易维护；同时刻意避开 #31731 的全局方案，保持改动面最小，是好的工程边界把握。

## 功能与动机

PR body 指出根因：GB300 PR 作业通过 bind(:0) 选择临时 nccl_port 并在释放后，rank 0 的 Torch TCPStore 再绑定同一端口时触发 EADDRINUSE，并附了两个失败链接（39539/39537 和 39541/39545）。第二次失败发生在 test_flashinfer_a2a.py，说明该竞争影响所有遗留 nccl_port=None 的 GB300 PR 服务器测试。

## 实现拆解

1. 在每个测试文件顶部从 DEFAULT_PORT_FOR_SRT_TEST_RUNNER 推导 NCCL_PORT_BASE（分别 +100/+110/+120），并注释保证端口低于 GB300 runner 临时端口下限 10240。
2. 为 test_deepseek_v3_cutedsl_4gpu.py 的 TestDeepseekR1Nvfp4CuteDSLDeepEP 和 TestDummyWithSBO 两个服务器启动添加 --nccl-port（BASE、BASE+1）。
3. 为 test_flashinfer_a2a.py 的三个测试类添加 --nccl-port（BASE、BASE+1、BASE+2）。
4. 为 test_disaggregation_aarch64.py 的 start_prefill 和 start_decode（并发启动）分别添加 BASE 和 BASE+1。
5. 配套验证：AST 检查确认七个启动都有显式端口、端口唯一且低于 10240；Ruff 0.15.1 lint/format、py_compile、pre-commit 全绿。

关键文件：

- test/registered/4-gpu-models/test_deepseek_v3_cutedsl_4gpu.py（模块 模型测试；类别 test；类型 test-coverage）：DeepSeek CuteDSL（NVFP4 + DeepEP）主线测试，两个服务器启动首次获得固定端口，覆盖了最初报告的失败场景。
- test/registered/ep/test_flashinfer_a2a.py（模块 EP 测试；类别 test；类型 test-coverage）：三个 FlashInfer A2A 测试类均受影响，是第二次失败点（test_flashinfer_a2a.py）所在文件。

- test/registered/disaggregation/test_disaggregation_aarch64.py（模块 分离部署；类别 test；类型 test-coverage）：AArch64 分离部署测试并发启动 prefill 和 decode 两个服务器，最能体现 " 每个启动独立端口 " 的设计约束。

关键符号：TestFlashinferA2ATrtllmRoutedFP4.setUpClass, TestFlashinferA2ACutedslStaticFP4.setUpClass, TestFlashinferA2ATrtllmRoutedFP8.setUpClass, TestDeepseekR1Nvfp4CuteDSLDeepEP.setUpClass, TestDummyWithSBO.setUpClass, TestDisaggregationMooncakeAARCH64Accuracy.start_prefill, TestDisaggregationMooncakeAARCH64Accuracy.start_decode

## 关键源码片段

### test/registered/ep/test_flashinfer_a2a.py

三个FlashInferA2A测试类均受影响，是第二次失败点（test_flashinfer_a2a.py）所在文件。

```
# 从测试工具共享的默认端口常量推导出本文件的端口基准。
# GB300 runner 的临时端口区间从 10240 开始，
# 因此把 NCCL 端口固定在其下方，避免 bind 竞争导致的 EADDRINUSE。
NCCL_PORT_BASE = DEFAULT_PORT_FOR_SRT_TEST_RUNNER + 110

class TestFlashinferA2ATrtllmRoutedFP4(CustomTestCase):
    """flashinfer A2A + flashinfer_trtllm_routed 的 modelopt_fp4 路径（DeepSeek V3）。"""

    @classmethod
    def setUpClass(cls):
        cls.model = DEEPSEEK_V3_FP4_MODEL
        cls.base_url = DEFAULT_URL_FOR_TEST
        # 每个测试类使用不同的端口偏移，确保同一文件内多个启动不冲突。
        cls.process = popen_launch_server(
            cls.model,
            cls.base_url,
            timeout=SERVER_LAUNCH_TIMEOUT,
            other_args=[
                "--tp", "4",
                "--ep", "4",
                "--dp", "4",
                "--enable-dp-attention",
                "--nccl-port", str(NCCL_PORT_BASE),
                "--moe-a2a-backend", "flashinfer",
                "--moe-runner-backend", "flashinfer_trtllm_routed",
                "--quantization", "modelopt_fp4",
                "--disable-flashinfer-autotune",
                "--model-loader-extra-config",
                '{"enable_multithread_load": true}',
            ],
        )
```

### test/registered/disaggregation/test_disaggregation_aarch64.py

AArch64 分离部署测试并发启动 prefill 和 decode 两个服务器，最能体现 " 每个启动独立端口 " 的设计约束。

```
# 每个 GB300 PR 测试文件使用独立的基准偏移，
# 避免并行运行的多个测试文件共享相同端口区间。
NCCL_PORT_BASE = DEFAULT_PORT_FOR_SRT_TEST_RUNNER + 120

class TestDisaggregationMooncakeAARCH64Accuracy(PDDisaggregationServerBase):
    @classmethod
    def setUpClass(cls):
        ...
        # 非阻塞启动 prefill 与 decode 两个服务器，
        # 它们必须使用不同的 NCCL 端口，否则 rendezvous 相互干扰。
        cls.start_prefill()
        cls.start_decode()
        ...

    @classmethod
    def start_prefill(cls):
        prefill_args = [
            "--trust-remote-code",
            "--disaggregation-mode", "prefill",
            "--disaggregation-bootstrap-port", cls.bootstrap_port,
            "--tp", "2",
            # prefill 服务器占用基准端口。
            "--nccl-port", str(NCCL_PORT_BASE),
        ]
        ...

    @classmethod
    def start_decode(cls):
        decode_args = [
            "--trust-remote-code",
            "--disaggregation-mode", "decode",
            "--disaggregation-bootstrap-port", cls.bootstrap_port,
            "--tp", "2",
            "--base-gpu-id", "2",
            # decode 服务器占用基准端口 + 1，避开 prefill。
            "--nccl-port", str(NCCL_PORT_BASE + 1),
        ]
        ...
```

## 评论区精华

无正式 review 评论。issue 评论里作者发起了两次 /rerun-test：第一次（Run #30616149775）仍失败，第二次（Run #30618506277）全部通过，验证了修复有效。作者还主动做了 AST 层面的端口分配断言和格式检查。

- GB300 测试重跑验证 (testing): 固定端口后测试稳定通过，修复有效。

## 风险与影响

- 风险：改动仅涉及 test/ 下三个文件，无生产代码风险。潜在风险： 1) 端口固定后，若同一 runner 上并行运行其他未固定端口的测试或服务，仍可能撞端口，但端口选择低于 ephemeral 范围、且 runner 按作业隔离，冲突概率低； 2) 若 DEFAULT_PORT_FOR_SRT_TEST_RUNNER 常量未来调整超过 10240 边界，端口会进入临时端口区间，需要配套更新注释与验证； 3) 这些端口只覆盖当前七个启动，未来新增 GB300 测试需沿用同一约定。
  - 影响：对 GB300（4×GPU）PR 测试的稳定性有直接改善，消除了 DeepSeek CuteDSL、FlashInfer A2A、AArch64 分离部署测试的间歇性 EADDRINUSE 失败。对最终用户和运行时无影响；对团队的影响是 CI 基建约定：GB300 测试应显式分配 --nccl-port。
  - 风险标记：仅测试变更，固定端口依赖常量边界，端口分配约定需随新增测试同步

## 关联脉络

- 暂无明显关联 PR