PR #21728 完整报告

sgl-project/sglang

[Fix] Update supported custom_mem_pool types for mooncake

合并时间：2026-03-31 11:18

原文链接：http://prhub.com.cn/sgl-project/sglang/pull/21728

执行摘要

此 PR 更新了 mooncake 模块中的自定义内存池类型常量，从 INTRA_NVLINK 改为 INTRA_NODE_NVLINK，以确保与先前 PR #18259 的变更对齐，避免配置不匹配导致的潜在错误。变更极小，风险低。

功能与动机

根据 PR body，动机是“Due to the change in https://github.com/sgl-project/sglang/pull/18259, SUPPORTED_MOONCAKE_CUSTOM_MEM_POOL_TYPES need to be changed accordingly.”，目的是将 custom_mem_pool_types 与 INTRA_NODE_NVLINK 对齐，以保持系统配置的一致性。

实现拆解

仅修改了一个文件：python/sglang/srt/disaggregation/mooncake/utils.py，具体改动如下：

- 将常量 SUPPORTED_MOONCAKE_CUSTOM_MEM_POOL_TYPES 从 ["NVLINK", "BAREX", "INTRA_NVLINK"] 更新为 ["NVLINK", "BAREX", "INTRA_NODE_NVLINK"]。 没有其他函数或模块变动。

评论区精华

无讨论或争议，只有 reviewer ShangmingCai 的批准，body 为空。

风险与影响

- 风险：改动极小，风险低，但若不更新可能导致 mooncake 内存池初始化时使用错误类型，引发运行时错误。具体风险点在 utils.py 文件中的常量使用路径。
- 影响：影响范围有限，仅涉及 mooncake 模块的内存池配置，确保与其他变更保持一致，避免系统不一致性。

关联脉络

与 PR #18259 直接关联，后者引入了 INTRA_NODE_NVLINK 的变化，需要此 PR 更新常量以匹配。从近期历史 PR 分析中未见其他直接相关 PR，但整体脉络指向维护硬件后端和内存池配置的一致性。