PR #6929 完整报告

PaddlePaddle/FastDeploy

[BugFix][KVCache] Fix mm hash boundary comparison in get_block_hash_extra_keys

合并时间：2026-03-30 17:13

原文链接：http://prhub.com.cn/PaddlePaddle/FastDeploy/pull/6929

执行摘要

- 一句话：修复 KVCache 中多模态 hash 边界比较的 off-by-one 错误，确保 prefix cache 计算正确性。
- 推荐动作：该 PR 值得工程师精读，特别是处理边界条件的逻辑和测试设计，有助于学习在类似场景中避免 off-by-one 错误。

功能与动机

根据 PR body，修复动机是当图像恰好与 block 相接（image_end == block_start）时，原代码使用 < 判断导致图像被误判为与 block 有重叠，影响 prefix cache 的 hash 计算正确性。

实现拆解

实现包括两个关键改动：在 fastdeploy/cache_manager/prefix_cache_manager.py 中，将 image_offset + image_length < start_idx 改为 <=（共两处，快速返回检查和循环内跳过检查）。在 tests/cache_manager/test_prefix_cache_manager.py 中，新增三个测试函数：test_get_block_hash_extra_keys_boundary_cases、test_get_block_hash_extra_keys_no_overlap_at_boundaries 和 test_get_block_hash_extra_keys_image_crosses_block_boundary，覆盖图像与 block 的各种位置关系。

关键文件：

- fastdeploy/cache_manager/prefix_cache_manager.py（模块 cache_manager）：包含核心逻辑修复，调整边界比较操作符，影响 hash 计算正确性。
- tests/cache_manager/test_prefix_cache_manager.py（模块 cache_manager）：新增边界条件测试，覆盖图像与 block 的各种位置关系，确保修复有效性。

关键符号：get_block_hash_extra_keys

评论区精华

review 中仅有批准，没有深入讨论。reviewer juncaipeng 直接批准，表明变更被接受无异议。

- Approval (other): 变更被接受

风险与影响

- 风险：风险较低：修改仅涉及比较操作符的边界条件，逻辑简单，但影响关键路径 get_block_hash_extra_keys，若未正确处理可能导致 cache 错误。新增测试覆盖了边界条件，增强了信心，兼容性方面是 bugfix，不引入 breaking change。
- 影响：对用户：修复自动生效，提高 prefix cache 的正确性，减少潜在错误。对系统：确保多模态输入在 cache 管理中的正确处理，提升可靠性。对团队：提供了边界条件的测试用例，可作为参考学习。
- 风险标记：边界条件修复，测试覆盖增强

关联脉络

- 暂无明显关联 PR