

dev2 直连 GPU 一存储：隔离测试与可行方案

结论

已完成隔离小测试：dev2 在内存中生成 256 KiB 合成数据，通过既有公网与 Tailnet 两条 SSH 路线直接写入 GPU 一的独立 SSD，再读回进行 SHA256 校验，两条路线均通过。dev2 未生成这份数据的磁盘副本。本轮未挂载远程文件系统、未迁移项目路径，也未改动开发任务。

记忆与现场核验

项目记忆 reference_gpu1_27b_machine_profile_2026-09-01.md 指向 gpu-prod，即运行 27B 模型的 GPU 一；现场主机标识为 704d4ca1274c。目标存储为 /cloud/cloud-ssd1，本轮实际剩余 211.7 GiB。/model 是共享模型库，不纳入本方案。先前核算的旧工作区和备份合计约 8.94 GiB，只是归档候选上限；未确认停用和保全独立改动前，不能当成已释放空间。

实际测试结果

测试数据 262144 字节，公网路线写入总耗时 5.360 秒、读回 4.078 秒；Tailnet 路线写入 6.217 秒、读回 1.678 秒。耗时包含 SSH 建连与认证，不能换算为稳定磁盘性能或大文件带宽。两次读回均与原数据完全一致，并核验实际位于目标 SSD。此前 4 MiB 测试在 30 秒内未完成，因此大文件速度与稳定性尚未通过。

对现有任务的保护

dev2 原有 6 个 Codex 进程在测试后仍存在。GPU 三项服务 PID 36、32、46851 的启动标识前后相同，8080、8081、8082 的健康检查均返回 HTTP 200。未重启任何服务，未修改现有项目、Git 或数据库路径。测试目录、合成数据、临时私钥与单条临时授权已清理。这些证据证明服务持续运行，但不等于已验证所有业务请求的性能完全无波动。

可行方案与下一步

若只要求报告、图片和备份直接存入 GPU，可给这些产物设置独立远端目录，通过受限上传入口直接写入；现有业务任务仍按原路径运行。若希望开发程序像本地文件一样直接读写 GPU 空间，需要远程挂载，并先验证大量小文件、锁、原子改名、断线恢复与速度。若要求连新工作区、缓存和中间文件也不占 dev2，则需将新任务的执行环境一起放到 GPU 一，并隔离 CPU、内存和磁盘资源；生产模型服务不在迁移范围。云机系统自身仍会产生日志。建议下一轮仍使用独立测试目录，先验证真实规模和断线行为，再决定是否切换新增产物路径。SSHFS 官方说明指出断线重连后旧文件句柄需要重开，参考：<https://github.com/libfuse/sshfs/blob/master/sshfs.rst>。本轮未实施永久挂载或正式迁移。