按项目阶段拆分计算、存储与结果迁出安排

完整操作教程 · 参考固定 · 往返恢复验收

两三个月的项目租服务器，最容易遗漏的不是核心数，而是数据什么时候进入、哪些中间结果必须保留，以及最后几天能否把结果取走。先按项目阶段拆分资源，才能让租期覆盖真实工作，并把计算费用和保存费用分开核对。本文给出从现有数据到迁出的计划方法，适合有明确任务和截止日期、准备按项目租用算力的研究者。

运行边界：示例代码需按安装版本和真实项目输入复核；未提供真实生物数据，语法或小样本检查不代表科学结果和性能验收。

两三个月的项目租服务器，最容易遗漏的不是核心数，而是数据什么时候进入、哪些中间结果必须保留，以及最后几天能否把结果取走。先按项目阶段拆分资源，才能让租期覆盖真实工作，并把计算费用和保存费用分开核对。本文给出从现有数据到迁出的计划方法，适合有明确任务和截止日期、准备按项目租用算力的研究者。

第一步：把项目拆成四个可交接阶段

不要把“用了三个月”直接理解为每天都需要最高配置。数据准备、正式计算、复查和结果迁出分别填写输入、输出、完成条件与负责人。只有当流程本身支持停机、恢复或环境迁移时，才能考虑分阶段调整资源；部分软件重建索引、重传数据库的代价很高，短暂停用未必节省总体成本。

阶段	要完成的工作	主要资源	离开这一阶段的证据
准备	数据校验、环境与参考验证	上传时间、数据库盘、少量计算	样例通过且输入清单冻结
计算	分批正式运行	CPU或GPU、内存、工作盘	任务成功、输出和日志齐全
复查	检查异常、补跑必要样本	可复用环境及适量计算	样本清单无未解释缺项
交接	下载结果、环境与日志	网络、保存空间与校验	接收方验证文件可用

第二步：分别核算四类存储

工作盘预算要同时考虑输入、索引或数据库、中间文件和最终结果。压缩FASTQ转换、BAM排序和组装等过程可能产生临时空间，不能仅把原始文件体积相加。以下示例只说明核算方式：输入200 GiB、参考80 GiB、预计临时峰值300 GiB、结果100 GiB，若它们在同一时刻都要保留，已是680 GiB；还需要结合文件系统可用量和后续增长安排空间。这里没有给任何具体流程承诺固定倍数。

```
# ■■■■■■■■■■■■■■
du -sh input reference work results
df -h .
df -i .
# ■■■■■■■■■■■inode■■■■■■■■■■■■■■■■
```

若输入和结果放在不同挂载点，要分别核算，不能把两个盘的剩余量当作一个盘来使用。网络盘、对象存储和本地工作盘的读写方式也不同；数据库是否允许长期保留、到期是否继续收费，应在开通前写明。备份要有独立副本和恢复安排，租用一块更大的工作盘不自动构成备份。

第三步：把传输时间纳入租期

先用少量非敏感样例测试实际上传和回传，而不是直接拿带宽标称值计算。线路、跨网访问、文件数量和校验计算都会影响传输时间。若临近到期才开始下载，剩余时间可能不够。上传后完成校验再进入正式计算，迁出时先取回不可替代结果和复现材料，临时文件按已确认的保留规则处理。

可以把整个安排写成下面的项目表。日期为示例，需按实际开始时间修改。最后的交接窗口应早于账号关闭，不应同时安排最重要的补跑与全部数据下载。

项目节点（示例）	交付记录	需确认事项
第1周准备	环境清单、输入校验、样例输出	软件版本和数据库是否可用
第2至第6周计算	逐样本状态与运行日志	资源是否允许按批次调整
第7周复查	缺项清单与补跑记录	异常样本是否需要额外租期
第8周交接	下载清单与接收方校验	到期保存、续租和删除规则

第四步：在报价单里确认变化与退出规则

至少问清初始资源、实际可用额度、租期起算点、附加存储、数据传输条件、环境部署范围和续租方式。若准备中途扩内存或更换GPU，还要确认是否会中断任务、是否换机器、数据和软件怎样迁移。不要把“支持扩容”理解为任何时刻都能无中断切换。

配套练习包

本PDF对应的ZIP包含本文代码块、主题练习模板或合成数据、README和SHA256校验文件。请从同一教程页面下载练习包；先核对校验值，再在独立目录复现。

参考资料

rsync官方手册 · 来源资料

sha256sum校验手册（Debian发布） · 来源资料

Nextflow缓存与恢复 · 来源资料

操作建议：先在合成小样本上验证流程，再依据项目参考、标签合同、存储策略与获配资源验收真实数据。