按软件和数据规模列出CPU、内存与磁盘需求

完整操作教程 · 参考固定 · 往返恢复验收

把“做生信”拆成软件、输入、并发与存储需求，形成能核对兼容性、可报价和可验收的服务器清单。

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。

配置服务器前，先把课题组未来一段时间要跑的分析拆成软件、输入、步骤和并发。CPU、内存和磁盘需求由最吃资源的步骤决定；只写“做转录组”或“需要大内存”，无法核对兼容性和可用空间。

先做任务清单

字段	填写内容	为什么要记
软件和版本	Cell Ranger 10.1、STAR等实际版本	硬件及系统支持随版本变化
输入规模	样本数、FASTQ大小、细胞数或基因组大小	决定任务是否可代表正式数据
分析步骤	比对、计数、整合、建模、可视化	不同步骤的峰值不同
并发和期限	同时运行数、完成日期	估算总资源和等待时间
存储	原始数据、参考、临时目录、结果、归档	可用空间不等于系统显示的总盘容量
限制	容器/调度额度、账号权限、数据规定	主机规格不等于单任务实际额度

用实际环境核对CPU、内存和磁盘

在候选Linux环境运行以下只读命令，将输出和任务清单一起保存。CPU型号之外还要看指令集；内存同时记录主机总量和作业额度；工作盘和临时盘分别查看挂载点与剩余空间。

```
lscpu
awk '/MemTotal|MemAvailable/ {print}' /proc/meminfo
grep Cpus_allowed_list /proc/self/status
df -h /path/to/work /path/to/tmp
ulimit -a
```

例如10x官方Cell Ranger v10.1系统要求列出至少8核、支持AVX+的Intel或AMD处理器、64 GB内存、1.5 TB可用磁盘和受支持的64位Linux；官方推荐值和不同规模数据集另有说明。它是该软件版本的基线，不代表所有生信软件，也不保证任意大样本在该配置下完成。应按正在使用的版本重新查系统要求。

把空间需求算到任务目录

磁盘清单至少分开统计输入、参考、临时文件、中间结果和最终结果。记录它们实际位于哪个文件系统；同一块盘上的“空闲空间”只能计算一次。可用 du -sh 查看现有目录大小，用 df -h 查看挂载点容量；任务运行中还要观察临时文件峰值，因为结束后清理会让峰值消失。

先标出未知项，再安排代表性试跑

把无法从文档确认的并发峰值、I/O或临时空间写成“待测”，不要用硬件型号推算。用接近正式流程的输入和参数跑一次，记录软件版本、输入规模、作业额度、步骤日志、退出状态和峰值来源。确认核心输出可读、维度或文件数合理、任务能在计划并发下复现后，再把实测结果用于配置比较。输入规模、软件版本或并发改变时，需重新验证。

做一张待确认项清单

建议把需求分成“已知、待测、需管理员确认”三类。例如软件官方最低内存是已知条件；代表性样本运行的峰值属于待测；容器内存上限、可用临时盘和CPU配额则要向平台管理员核对。每个数字都带单位、来源和日期，避免把GB与GiB、整机总量与单作业额度混写。并发规划时先写出每天有多少人会提交任务，再确认它们是否能同时运行；同机多个任务要按重叠时段估算，不能把单任务资源直接当作总需求。

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

10x Genomics Cell Ranger v10.1系统要求 · 来源资料

Linux proc文件系统 · 来源资料

Linux cgroup v2资源控制 · 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。