把可迁移流程搬到租赁服务器并核对结果

完整操作教程 · 参考固定 · 往返恢复验收

决定把一个可迁移步骤搬到租赁服务器之后，下一步是证明它在新环境里仍然完成了同一件事。直接复制一段sbatch脚本通常不够：模块、参考路径、工作目录和调度资源都可能变化。本文从一个已有、结果正确的校内任务出发，逐步打包环境说明、复制受控样例、重写启动方式并比较关键输出，不涉及改变分析方法。

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。

决定把一个可迁移步骤搬到租赁服务器之后，下一步是证明它在新环境里仍然完成了同一件事。直接复制一段sbatch脚本通常不够：模块、参考路径、工作目录和调度资源都可能变化。本文从一个已有、结果正确的校内任务出发，逐步打包环境说明、复制受控样例、重写启动方式并比较关键输出，不涉及改变分析方法。

冻结一个可作为基准的任务

选一个已经跑通的代表性样本，保存输入校验、命令、版本、参考说明、线程与随机设置、输出和日志。不要把一个尚未解释错误的任务当迁移基准，否则新机器上的差异无法归因。输入是否允许离开原系统先确认；如果只能使用公开或模拟数据验收，就明确它只能覆盖环境与流程，不能覆盖所有正式数据的规模。

迁移包内容	最小要求	常见遗漏
输入与参考清单	文件名、校验值、参考版本	只记目录，没有记实际版本
运行入口	可从项目根目录启动的脚本	依赖交互shell里的变量
环境说明	工具、R/Python包、容器或模块版本	以为module名字在新系统也存在
基准输出	必要文件与关键验收指标	仅保留一张最终图片
运行记录	退出码、日志、峰值、耗时	把排队等待和实际计算混在一起

把调度包装与分析命令分开

Slurm脚本中的#SBATCH参数由调度器解释，在普通shell里只是注释；相反，srun和module命令可能在新机器根本不存在。先提取真正执行分析的命令，整理成run_sample.sh，再由各系统各自的启动包装调用。保留输入参数和资源设置，不要在迁移时顺便更换方法或升级全部软件。

```
#!/usr/bin/env bash
# run_sample.sh■■■■■■■analysis.R■■■■■■■■■■■■■■
set -euo pipefail
sample_id="$1"
input_file="$2"
mkdir -p "results/$sample_id" "logs/$sample_id"
Rscript --vanilla analysis.R \
  "$input_file" "results/$sample_id" \
  > "logs/$sample_id/stdout.txt" \
  2> "logs/$sample_id/stderr.txt"
# analysis.R■■■■■■■■■■■■■■■■
```

检查脚本是否使用绝对路径、符号链接或临时目录，确认参考文件不再隐含指向校内共享盘。容器可以固定部分用户空间依赖，但仍要核对宿主CPU、驱动、挂载和权限；容器标签也可能变化，正式复现最好记录实际镜像摘要及来源。对不能迁移的许可和数据访问依赖，保留在原环境或改用获准方案。

复制样例并验证输入未变

下面使用示例主机名，不包含真实地址和账号。先核验目标主机身份，复制到新的项目目录；不使用删除目标内容的参数。rsync默认快速比较不是内容校验的替代品，重要输入仍需要核对校验清单。项目结构应避免同名文件互相覆盖。

```
# ■■■■■■■■■■■■■■■■■■■■■■■■
sha256sum input/pilot.tsv reference/annotation.tsv > inputs.sha256
rsync -av --partial ./ researcher@server.example.org:projects/pilot-migration/
# ■■■■■■■■■■■■■■■■■
sha256sum -c inputs.sha256
Rscript --version
Rscript --vanilla -e 'sessionInfo()' > environment-target.txt
```

在新机器运行，再比较正确的层级

用相同输入、参数和线程设置运行一次，保存退出码和资源日志。逐项核对输出文件是否齐全、样本标识是否一致、记录数和维度是否符合基准。确定性文本输出可比较内容；带时间戳的报告、压缩文件、随机模型及浮点计算不应盲目要求整个文件hash相同，应预先定义需要相同的字段或允许的数值偏差，并解释依据。

若结果不同，依次检查输入校验、参考版本、环境依赖、参数、随机种子、排序规则和执行是否完整。先修复这些差异，再比较性能。若任务报Killed，核对进程和账号限额；若输出缺一半，检查错误日志及磁盘，而不是把新机器速度更快视为成功证据。

扩大到批次前完成一次回传

把样例结果及运行材料取回原环境，再确认原下游步骤能读取。然后选择适量正式样本验证吞吐量和峰值，最后才制定整批并发。使用Nextflow的项目还应按官方规则处理缓存、work目录和输入路径；迁移输入路径会影响默认缓存判断，不能承诺复制目录即可无条件resume。

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

Slurm sbatch官方手册 · 来源资料

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

Docker镜像与容器运行 · 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。