盘点Arrow、中间文件和临时目录的空间占用

完整操作教程 · 参考固定 · 往返恢复验收

先定位Arrow、R临时目录和项目依赖，区分同盘峰值、配额与半成品。

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。

ArchR创建Arrow时磁盘不足，先定位写满的是Arrow输出、R临时目录还是QC和其他中间文件。Arrow是后续项目的重要数据，不是可随手删除的缓存。文件压缩后的大小也不能代表创建过程峰值，多个样本并行还会同时产生临时占用。

保留失败日志和实际路径

保存ArchR版本、createArrowFiles实际调用、输入片段或BAM清单、样本名与错误全文。当前接口的outputNames决定Arrow文件前缀，QCDir决定QC目录，cleanTmp控制工具临时清理；它们不能替代对工作目录和R临时目录的确认。旧教程filterTSS/filterFrags不要不加核对套给新版本。

位置	怎样定位	不能直接怎样做
Arrow文件	outputNames和实际生成路径	创建未完成就当有效Arrow
R临时目录	本次进程tempdir()	只改终端当前目录就假定转移
QC与日志	QCDir、ArchRLogs与调用记录	没保存错误就清空现场
项目引用数据	getArrowFiles返回的路径	当通用缓存批量删除
文件系统额度	df、inode、用户配额	把整盘剩余当账号可用量

已有项目时，导出Arrow依赖清单

以下针对曾用saveArchRProject保存的可信项目，不用于把半成品Arrow强行装成项目。保存为inventory_archr.R，传入项目目录；它只加载项目结构并列路径和文件大小，不读取整个片段矩阵。输出清单用于识别必须保留的依赖。

```
library(ArchR)
args <- commandArgs(trailingOnly=TRUE)
stopifnot(length(args)==1L, !file.exists("arrow-inventory.tsv"))
proj <- loadArchRProject(path=args[1], force=FALSE, showLogo=FALSE)
files <- getArrowFiles(proj)
print(packageVersion("ArchR"))
print(list(working_directory=getwd(), R_tempdir=tempdir()))
info <- data.frame(path=unname(files), exists=file.exists(files),
                   bytes=unname(file.info(files)$size))
write.table(info, "arrow-inventory.tsv", sep="\t", row.names=FALSE, quote=TRUE)
writeLines(capture.output(sessionInfo()), "archr-inventory-session.txt")
```

若项目加载因依赖缺失失败，先按原目录清单找回文件，不用force忽略问题来宣称恢复成功。Arrow还没建成、没有项目时，以输入调用和生成文件列表盘点即可。清单中的NA或exists为假是路径故障信号，不说明那个样本不需要保留。

把目录映射到真实存储

在Shell对实际输入、输出及R临时目录分别执行df -h、df -i和du -sh。df报告文件系统，du报告目录占用；几个目录同属一个文件系统时合并核算。若df有空间仍报错，进一步核对配额、权限和具体写入目标，而不是立刻重试更高线程。

需要改变R临时目录时，应在启动新R进程之前设置TMPDIR，并确保目录已经存在且可写。会话启动后才Sys.setenv，不能让已经创建的tempdir自动迁过去。不要把大临时文件放入未核对额度的内存盘，避免把磁盘不足变成内存终止。

```
# ■■■■■■■■■■■■■■■■■■■■■■
R_TMP=/absolute/scratch/archr-run-001
test -d "$R_TMP" && test -w "$R_TMP" || exit 2
TMPDIR="$R_TMP" Rscript --vanilla -e 'cat(tempdir(), "\n")'
```

修复容量后用新任务验收

先确认失败进程与相关写入都已停止，记录半成品和剩余文件，不清空共享临时目录。可以降低同时创建Arrow的样本数，再按同一输入与科学参数试跑；改minTSS、minFrags或矩阵生成选项会改变保留细胞及后续内容，不能只作为不影响结果的省盘技巧。

正式完成后，核对每份Arrow的样本身份、细胞范围、所需矩阵和QC结果，并确认项目能访问其全部依赖。只有文件存在或占用较小不算成功。将输入、Arrow、QC、临时峰值及同盘任务记在一张表中，下一批才知道是调整并发还是确实需要更多工作盘。

文件与目录检查使用实际运行账号，管理员能读写不证明当前作业账号具备同样权限。

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

ArchR createArrowFiles · 来源资料

ArchR getArrowFiles · 来源资料

ArchR loadArchRProject · 来源资料

R临时目录与TMPDIR · 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。