从对象大小、临时复制和可用内存定位分配失败

完整操作教程 · 参考固定 · 往返恢复验收

从对象尺寸、稀疏表示、复制行为与实际任务额度四个层面排查R内存分配失败，再判断是否需要大内存服务器。

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。

R提示 cannot allocate vector，通常表示当前R进程无法取得所需的连续对象空间；终端只显示 Killed，则可能是操作系统、容器或作业调度器终止了进程。两类现象的排查证据不同。先保存报错、最后完成步骤和资源记录，再决定是否调整对象、并行数或内存。

先分清R错误、并行限制和系统终止

现象	优先检查	不要直接得出的结论
cannot allocate vector of size	R进程位数、当前对象、内存额度、是否复制或稠密化	不能仅凭报错判断物理内存不足
future提示globals超限	worker要导出的全局对象及future配置	不是操作系统内存上限
终端显示Killed或退出码137	作业记录、容器/cgroup事件和系统日志	退出码本身不能证明OOM

在R会话中记录对象维度、近似大小、垃圾回收状态、版本和并行设置。以下代码假设对象名是 obj；把它替换成实际对象名。object.size()用于估计对象自身占用，不包含所有关联环境、外部指针或函数执行时的临时副本。

```
stopifnot(exists("obj"))
dim(obj)
format(object.size(obj), units = "GiB")
gc()
R.version.string
sessionInfo()
Sys.getenv(c("R_MAX_VSIZE", "R_FUTURE_GLOBALS_MAXSIZE"))
```

若使用future，future.globals.maxSize限制的是发送到worker的全局变量大小，不等同于RAM额度。盲目调大它只会允许更大的对象被导出；若多个worker各自复制对象，实际内存可能更高。先记录worker数量和失败步骤，再用单worker复跑小样本判断是否由并行副本触发。

估算一次稠密化需要的最低载荷

矩阵若从稀疏格式转为双精度稠密格式，单个数值矩阵的理论载荷约为“行数 × 列数 × 8字节”。例如2万个基因、10万个细胞，单个稠密矩阵约14.9 GiB；这还未包括原始对象、索引、名称、临时副本和R本身。下面只是算术估算，不要为了测大小而对大型对象调用 as.matrix()。

```
genes <- 20000
cells <- 100000
dense_gib <- genes * cells * 8 / 1024^3
print(dense_gib)
# ■■■■■■■■■■■■
```

检查报错所在函数是否把稀疏矩阵转为稠密、是否同时保留旧对象与新对象、是否有多个worker复制输入。不要未经科学判断删除基因、样本、assay或layer来“修复”内存问题；这可能改变研究问题或结果。

检查系统额度并做单变量复跑

Linux终端可先记录主机可用内存、当前shell的虚拟内存限制和作业环境。共享服务器上主机总内存不等于当前账号或容器可用额度。若任务运行在Slurm等调度器内，优先查看该作业的请求值、实际峰值、退出状态和调度器事件；没有权限读取系统日志时，请管理员按作业ID核对。

```
date -Is
free -h
ulimit -a
# ■Slurm■■■■■■■■■■JOB_ID
sacct -j JOB_ID --format=JobID,State,ExitCode,ReqMem,MaxRSS,Elapsed
```

一次只改一个因素：先把worker降到1；若仍失败，再检查数据结构和失败步骤；确认触及作业内存上限后，才用相同输入评估更高额度。每次用独立输出目录，记录代码版本、R与包版本、输入规模、并行数、退出码和峰值口径。成功标准是目标步骤完整结束、输出维度及关键字段符合预期，并能在约定的并发条件下复现，而不是只看到进程没有报错。

形成可复查的结论

将错误原文、最后成功步骤、对象维度、并行设置、资源额度和复跑结果放在同一记录中。如果R报告无法分配但系统额度未触顶，优先排查单个对象过大、连续空间、稠密化或对象复制；如果系统事件显示内存限额触发，再比较减少并行与提高额度的效果。R的堆上限变量只影响R可使用的对象空间，不会增加机器实际内存。版本、输入规模或工作流变化后，应重新验证原结论。

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

R Memory Limits · 来源资料

R object.size · 来源资料

future options · 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。