核对索引与范围读取，演练远程查看和访问控制

完整操作教程 · 参考固定 · 往返恢复验收

检查HTTP206、Content-Range及索引，再分别验收桌面IGV、igv.js与访问权限。

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。

大BAM或bigWig可以按显示区域远程读取，前提是格式、索引、参考与HTTP范围请求都正确。把文件放进网页目录并给一个下载链接，还不等于IGV可用。本篇先用无敏感内容的测试文件检查范围响应，再验收实际客户端的索引和受控访问。

先对齐数据与客户端要求

坐标排序BAM需要匹配且受客户端支持的索引，如BAI；bigWig内部有索引，但同样需要服务端支持范围读取。CRAM还涉及匹配参考及CRAI。记录参考组装、contig名称与文件指纹，先在本地或已挂载路径用目标IGV版本验证代表区域，避免把参考不一致误当网络问题。

对象	本例约定	验收要点
数据	稳定URL及正确文件格式	登录页HTML不是真实数据响应
索引	匹配数据的索引URL	权限和有效期也需覆盖索引
HTTP	206和正确Content-Range	不能仅看Accept-Ranges头
客户端	桌面IGV或igv.js分别验收	浏览器跨域还涉及CORS

用64字节请求检查服务器

将URL换成你有权访问、长度超过64字节的非敏感测试文件。需要curl8.4以上，限制返回大小和等待时间；本例不使用自动HTTP解压。真实项目按获批认证方式接入，不能把项目访问凭据写到公开文章、URL或诊断输出中。

```
#!/usr/bin/env bash
set -euo pipefail
url=https://data.example.org/test/test.bam
mkdir range-check-001
root="$PWD/range-check-001"
curl --version > "$root/curl-version.txt"
python3 - "$root/curl-version.txt" <<'PY'
import re,sys
text=open(sys.argv[1]).read()
m=re.match(r'curl ([0-9]+)\.([0-9]+)\.([0-9]+)',text)
assert m and tuple(map(int,m.groups()))>=(8,4,0)
PY
status="$(curl --silent --show-error --fail --location \
  --range 0-63 --max-filesize 64 --max-time 15 \
  --header 'Accept-Encoding: identity' --dump-header "$root/headers.txt" \
  --output "$root/range.bin" --write-out '%{http_code}' "$url")"
test "$status" = 206
python3 - "$root" <<'PY'
import re,sys
from pathlib import Path
root=Path(sys.argv[1]);headers={}
for line in (root/'headers.txt').read_text().splitlines():
    if line.startswith('HTTP/'):headers={}
    elif ':' in line:
        key,value=line.split(':',1);headers[key.strip().lower()]=value.strip()
m=re.fullmatch(r'bytes 0-63/([0-9]+)',headers.get('content-range',''))
assert m and int(m.group(1))>=64
assert headers.get('content-encoding','identity').lower()=='identity'
assert (root/'range.bin').stat().st_size==64
print('range response verified')
PY
```

仅声明Accept-Ranges不等于真的按范围返回，HTTP200整文件响应也不能当作通过。反向代理、登录跳转、缓存或动态压缩可能改变行为，分别查实际最终响应。这个检查只覆盖一个范围，不证明索引、认证续期或整个数据内容均正确。

在目标IGV里验证实际浏览

桌面IGV按File里的Load from URL输入数据URL，并按界面提供索引URL，选对参考后查看几个已知区域。核对覆盖、样本和信号形态，再跨远距离区域跳转，观察日志与实际传输。有效空白区域和无法加载要分开，不能只凭第一个截图就判定全部正常。

igv.js运行在浏览器中，跨源访问需配置相应CORS，且与认证方式兼容；桌面IGV不应被套用同一浏览器跨域判断。数据、索引、参考以及必要的重定向都要在同一权限方案下验收。临时URL过期后刷新和重新加载的行为也应测试，而不是无限延长所有链接。

按访问量测网络与存储

按真实使用人数、区域大小、随机跳转和网络链路测响应与传输，范围读取仍会消耗网络、索引和服务资源。数据更新后同步索引、指纹和缓存策略，不能保留旧索引继续指向新BAM。原文件可恢复副本、访问控制、到期和日志按项目安排，远程浏览不是备份。

只有持续计算和查看确有需求，并且候选访问方案已经验证，才据此评估服务器与存储。一次curl成功不代表所有学生或合作方的校园网、客户端和权限都可用，正式交接应由实际用户完成一次加载与区域检查。

保留不含敏感信息的测试响应和检查时间，便于区分代理配置变化、索引替换与数据更新。

HTTP 远程查看不是“把文件放上网”

IGV 远程查看适合浏览需要随机访问的比对文件或信号轨道，不等同于把整个 BAM 下载到浏览器，也不能只凭一个文件 URL 判断服务可用。BAM 通常需要对应的索引文件，BAM 与 BAI 必须来自同一份排序、索引后的数据；CRAM 还依赖对应参考序列和索引，参考序列版本或 contig 命名不一致可能表现为轨道空白。bigWig 通常自带索引结构，但服务器仍需支持客户端按范围读取。先确认文件格式、索引文件名和数据版本，再排查网络层问题。

IGV 桌面版与 igv.js 浏览器版的服务要求略有区别。浏览器内的 igv.js 通常会受跨域资源共享（CORS）限制，服务端需要允许实际网页来源，并支持 HTTP Range；桌面版没有网页同源策略，但仍需要可达的 URL、正确的响应头和访问权限。不要把某一种客户端的测试结果直接推广到另一种客户端。若要嵌入网页，先在目标域名和目标浏览器上验证，再在同一环境内用 IGV 桌面版独立验证，区分是数据服务器、浏览器策略还是客户端配置问题。

正确理解 Range 响应

客户端随机读取一个区域时，通常会发送带 Range 的 HTTP 请求。成功返回部分内容时，常见响应状态为 206，并带有与请求范围对应的 Content-Range。仅看到响应头里有 Accept-Ranges，不能证明代理链、对象存储和权限网关都正确实现了范围读取；有些服务器会忽略请求并返回整文件的 200，也有的代理在压缩或重定向后改变响应行为。建议用小范围请求检查状态码、Content-Range、Content-Encoding 和实际返回字节数，并从真实的 IGV 客户端再做一次区域跳转验证。

测试要覆盖文件开头和非零偏移的范围，因为某些对象存储或代理只在特定区段出错。测试用公开的非敏感样例，避免把真实项目文件的 URL、访问令牌、样本名或研究对象写入公共工单、命令历史和文章截图。不要用一次成功的 64 字节请求就宣称大文件远程浏览正常；至少加载一个已知坐标区域，来回切换染色体与缩放级别，确认索引访问和内容响应都正常。

权限、凭证与数据安全

科研数据可能包含未发表结果、个人信息或受合作协议约束的材料。公开 URL 不是默认合适的交付方式。优先使用学校认可的 VPN、登录网关或带有效期限的签名 URL，并用最小权限账号只读开放需要的文件。签名链接过期时间、可访问对象范围和日志留存都要与课题组的伦理、数据管理计划及合作方约定一致。不要把长期密钥写进 IGV session 文件、网页源代码、公共脚本或博客；需要演示时，使用人工生成或已获许可的去标识数据。

反向代理与对象存储之间还需要检查身份传递：请求能访问 BAM 却访问不到 BAI，或数据文件有权限而索引文件没有权限，都会造成客户端报错。访问日志不应记录完整的敏感查询参数；若服务使用带令牌的 URL，需了解服务器、代理和浏览器会不会把完整 URL 写入日志。课程材料或公开教程中的示例域名只能作为占位符，真实部署时应替换成有访问控制的地址，不要让读者误以为把文件放到任意公共 Web 目录就完成了安全共享。

按顺序定位“轨道不显示”

第一步，在服务器端确认文件和索引存在、路径大小写正确、文件权限允许目标服务读取。第二步，确认链接重定向后的最终地址仍可访问，响应的 Content-Type 与文件类型相符，没有登录 HTML 页面伪装成成功响应。第三步，分别对 BAM 和索引请求做范围读取，检查状态码与范围。第四步，检查 BAM 与索引是否是同一版本，并用 samtools quickcheck 等工具确认 BAM 结构和头部可读取；quickcheck 只做有限的文件结构检查，不能替代完整数据审查。第五步，在 IGV 中加载明确的参考基因组，查看染色体名称和文件头是否一致。

浏览器端还要检查开发者工具中的网络错误和 CORS 预检请求。允许来源应限制为需要使用的站点，避免为了“先让它能显示”而配置任意来源。对于需要认证的请求，跨域凭证、允许的请求头和浏览器策略必须一起配置；CORS 不是访问控制本身，也不应被当成身份验证替代品。如果某种访问控制与客户端的 Range 请求不兼容，应先设计安全的网关方式，再进行实际客户端测试。

选择合适的交付方式

若课题组只有少量固定用户、数据受控且只需定期浏览，可以在校内存储与 VPN 后提供受控访问；如果需要在网页中为多人嵌入轨道，需评估 CORS、并发、访问日志和授权回收流程。文件很大、用户分散或网络质量不稳定时，可考虑由学校认可的对象存储或受控计算平台托管，但应验证其 Range、认证与生命周期策略。下载后本地查看在一些场景更简单，但会带来副本管理、磁盘空间和数据外流风险；要根据项目的数据治理要求选择，而不是假设远程加载天然更安全。

试运行时记录客户端版本、参考基因组版本、数据文件校验和、访问方式、测试区域、响应状态和失败现象。将可复现的最小样例交给管理员排查，样例不能包含真实受限数据。这样能够把“IGV 看不到”分解为文件格式与索引、HTTP 范围、身份权限、跨域策略、参考版本和客户端配置几个可验证的问题，也能避免盲目增加带宽或服务器配置。

什么时候才需要增加服务器资源

单个远程浏览请求的主要瓶颈可能是网络往返延迟、对象存储读取性能、索引是否可达、代理缓存策略或客户端并发，而不一定是 CPU 或内存。先用实际用户的网络和目标区域测量首屏时间、连续跳转耗时、服务端响应字节数和并发访问情况。若服务器没有压力而用户仍等待，升级 CPU 往往无效；若多个课题组同时访问同一数据集，可评估只读缓存、对象存储吞吐和访问限流，但必须确保缓存不绕过权限检查，也不把私有数据留在公共缓存中。

给课题组做一次可复现验收

部署验收可以准备一份无敏感信息的测试 BAM 及索引、一个已知坐标区域和一台校外网络客户端。先记录文件大小与校验和，确认浏览器或 IGV 桌面版的访问方式，再依次测试文件链接、索引链接、非零偏移 Range、目标区域加载、缩放与跳转、过期凭证后的拒绝行为。测试结果写明使用的客户端版本、网络位置、服务端响应状态和错误信息；不要只截一张能显示轨道的图作为验收依据。访问控制也应测试未授权用户确实无法读取，而不是只验证授权账号能打开。

签名 URL 需要注意分享和撤销：通过邮件、聊天、会话文件或截图转发后，可能超出原定用户范围。若业务要求多人长期查看，使用可管理的账号和权限组通常比长期有效的公共链接更易审计。定期抽查过期机制、权限回收和访问日志，并确

保缓存策略不会在授权撤销后继续返回私有数据。若无法满足这些要求，就不应把真实项目的 BAM 发布到公开网页；可改用校内受控访问或本地分析方式。

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

IGV桌面版官方URL加载与范围要求 · 来源资料

igv.js数据服务器与CORS要求 · 来源资料

curl官方Range与大小限制 · 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。