

GigaScience

通过提高传输速度，帮助生命科学社区提高研究数据的可访问性

GigaScience 是一家支持开放式访问和开放式数据的在线生命科学期刊，由 BGI 和 BioMed 联合出版。该期刊发表的“大数据”文章覆盖生物科学和生物医学科学的各个领域，包括基于很难访问的数据的学科，如影像学、神经科学和系统生物学。期刊接收和发表的所有文稿都聚焦大型数据集的使用、分析或工具开发。

GigaScience 发现再现数据密集型科学研究存在问题，于是，他们打算提供一个解决方案。他们的目标是让研究变得可再现和可重复使用，让研究文章变得透明，让大型数据便于访问和引用。为了实现这一目标，GigaScience 在一个全面的公共数据库 GigaDB 中托管了与每篇发表的文章相关的所有数据集。它为每个数据集提供了一个“数字化对象标识符”。这样，人们就能更轻松地定位他们所找的文件；此外，人们在重新使用或再现研究时，能够直接引用数据。

为了处理如此庞大的数据集传输任务，GigaScience 采用了 IBM Aspera 软件产品套件，为作者、审校者和其他用户提供工具，以最快的速度上传和下载文稿配套的所有大型数据集。

更快速地进行进一步传输 TB 级别的研究数据

为支持 GigaScience 期刊发表的文章而提交的数据集可能规模高达数 TB。GigaScience 发现 FTP 不太适合用于传输大型文件，因为传输速度往往太慢，一旦用户碰到网络问题，传输就必须从头开始。此外，由于网络的高度延迟，远距离传输相当耗时，也很不可靠。

有一次，GigaScience 在上传 15TB 的肝癌数据集时面临挑战¹。排除了 FTP 后，GigaScience 不得不将数据加载到 8 个硬盘驱动器上，以物理方式将数据从提交者处运送至期刊总部，这是一个耗钱耗时的流程。

GigaScience 和研究人员无法等待数周的时间来完成文稿配套的大型数据集的上传。期刊通常希望在两周内将审校意见返回给作者，并在作者接收意见后立即发表文章；而研究人员需要快速下载文章，这是最有效的运转方式。GigaScience 希望寻找一款传输解决方案，来为作者、审校者和内容使用者提供一个快速、高效的机制，实现 GigaDB 上大型数据集的上传和下载，同时支持技能水平各不相同的用户轻松使用数据。

概述

挑战

将大型数据集从配套文稿上传到 GigaScience 数据库 GigaDB，或从 GigaDB 下载大型数据集。

业务转型

内容提交者和审校者及研究用户使用 IBM® Aspera® Connect 插件，以最快的速度上传和下载大型数据集。

成果

上传大型数据集只需几小时而非几天。

借助高速传输功能，GigaScience 能够更快速地审校、接收和发表文稿，并在两周内将决策返回给提交者。



*“ 我们想要一款
面向广大用户的
解决方案，而非
仅面向有技术专
业知识的用户。
通过使用
Aspera，任何人
都能轻松上传或
下载数据。
Aspera 是目前
唯一一款能帮助
人们及时访问数
据的解决方案 ”*

---Laurie Goodman 博
士，GigaScience 总编
辑

在作者、审校者和读者之间快速共享数据

在评估了多个选项后，GigaScience 选择利用 IBM® Aspera® Connect Server，将提交的文稿配套的所有数据集快速传输到 GigaScience 数据集；利用 IBM® Aspera® Console，管理和监控整个端到端传输流程。作者使用 Aspera 免费下载 Aspera Connect 插件，将文稿相关的数据集提交到 GigaScience 的私有数据存储站点。然后，审校人员访问这些文件，使用浏览器插件高速下载和上传文件。如果期刊决定发表文章，那么数据将通过 Aspera 传输到期刊的公共数据库 GigaDB，让期刊读者能够使用 Aspera Connect 插件，轻松查看和下载这些数据。

解决方案组件

- IBM® Aspera® Connect Server
 - IBM® Aspera® Connect 浏览器插件
 - IBM® Aspera® Console
 - IBM® Aspera® Cargo
-

更好地访问和再现研究和文章

过去，用户需要花一整周的时间来下载少量数据。借助 Aspera，不论数据集规模如何，用户都能快速、可靠、轻松地将数据上传到 GigaDB 以及从 GigaDB 下载数据。

“人们希望使用这些数据；他们不希望花一周时间坐等下载数据，” GigaScience 总编辑 Laurie Goodman 说道，“Aspera 是目前唯一一款能满足期刊要求、帮助人们及时访问数据的解决方案。”

借助 Aspera，GigaScience 为期刊读者提供了更快速、更可靠、更愉悦的下载体验。GigaScience 获得了用户的大加赞赏。其中一位用户用了几个小时而非几天就完全下载了 1.2TB 的数据集，相比以前使用 FTP 时速度有了大幅改进，这一点给该用户留下了深刻的印象。

此外，不论用户的计算能力如何，Aspera 都能一视同仁地支持精通计算机的用户和非技术用户上传和下载数据。“我们想要一款面向广大用户的解决方案，而非仅面向有技术专业知识的用户，”Goodman 补充道，“通过使用 Aspera，任何人都能轻松上传或下载数据。”

借助 Aspera 的速度、可靠性和便捷性，GigaScience 能够实现帮助生命科学社区更好地访问和再现研究与文章的目标。

Aspera 软件的其他显著优势还包括：

- 快速传输：使用 Aspera Connect Server，以最快的速度将数据上传到 GigaDB 以及从 GigaDB 下载数据，不论文档有多大，传输距离有多远，或者网络情况如何。
- 易于使用：借助直观的 Web 界面和自安装 Aspera Connect 插件，Aspera 支持每位 GigaDB 用户轻松上传和下载数据，不论他们的计算机使用能力如何。
- 可靠性：通过自动恢复和重试部分传输或失败的传输任务，GigaScience 及其用户确信他们可靠地完成了传输。

了解更多信息

如欲了解有关 IBM Aspera 产品如何为您提供帮助的更多信息，请联系您的 IBM 代表或 IBM 业务合作伙伴，或访问以下网站：

ibm.com/cn-zh/cloud/high-speed-data-transfer

为何选择 IBM 云计算？

采用云计算，为竞争优势加上智胜砝码。IBM 云可确保无缝地集成到公共和私有云环境。基础架构安全、可扩展而且灵活，可提供定制的企业解决方案，这些都使 IBM 云成为混合云市场的领导者。

立刻免费咨询 IBM 专家:400 668 0529（工作日 9:00-17:00），敬请访问：ibm.com/cn-zh/cloud



关注 IBM 云公众号



访问 IBM 云官网



了解 IBM Aspera



© Copyright IBM Corporation 2019

IBM Corporation

中国印刷

2019 年 9 月

IBM、IBM 徽标、ibm.com、及 IBM Aspera 是 International Business Machines Corporation 在世界各地司法辖区的注册商标。Web 站点 ibm.com/legal/copytrade.shtml 上的“Copyright and trademark information”部分中包含了 IBM 商标的最新列表。

其他公司、产品或服务名称可能是其他公司的商标或服务标记。本文档中所含信息仅供参考之用。尽管出版时尽力对本文档所含信息的完整性和准确性进行了验证，但所有信息均按原样提供，不对其作出任何明示或暗示的保证。此外，此类信息基于 IBM 的当前产品计划和策略，如有更改，恕不另行通知。由于使用本文档或其他材料或由于其相关事宜而造成的损害，IBM 不负任何责任。本文档中的任何信息均不表示 IBM 或其供应商或许可方作出任何保证或陈述，也不会更改对 IBM 软件的使用具有约束力的条件和条款。

本案例研究说明了所述 IBM 客户使用 IBM 产品的方式。我们不保证您的使用会出现相当的结果。在本出版物中，但凡提及 IBM 产品或服务时，并不表示 IBM 可以在 IBM 业务所涉及的所有国家或地区予以提供。