

基因组调控中心 / 生物医学和基因组学研究中心

通过高速传输文件，支持前沿研究

概述

挑战

基因组调控中心 / 生物医学和基因组学研究中心, Centre de Regulació Genòmica (CRG) 希望帮助全球研究机构共享基因测序数据。怎样才能快速、经济高效地传输如此庞大的数据呢？

业务转型

借助高速的 IBM® Aspera® 文件传输平台, CRG 最高能够以 3 Gbit/s 的速度共享文件, 让研究人员能够快速访问所需的数据, 用于实施前沿基因组项目。

成果

最高 500

最高可轻松打包 500TB 的离散序列数据用于共享。

最高 3

传输速度最高可达到 3 Gbit/s。

支持

支持前沿基因组研究。

生物医学和基因组学研究中心的基因组调控中心, Centre de Regulació Genòmica (CRG, 英文名为 the Centre for Genomic Regulation) 成立于 2000 年, 是一家知名的生物医学研究机构, 位于巴塞罗那。该机构拥有超过 600 名员工, 目标是推动知识的增长, 为社会、公共卫生和经济的繁荣发展作出贡献。

向全球受众呈现基因数据

全球各地的研究人员致力于发现个体基因组成的差异如何影响人类的健康。随着基因测序技术变得越来越快、越来越便宜、越来越准确, 可用于分析的数据量也在迅速增长。研究机构如何快速、便捷且经济高效地共享这些数据？

CRG 机构 IT 总监 David Camargo 讲述了整个故事: “尽管基因测序成本不断降低, 但是收集数据用于疾病的检测、预防和治疗依然需要我们投入大量时间和资源。”

“为了解决这一挑战, CRG 与欧洲生物信息研究所 [EMBL-EBI] 合作改进了欧洲基因组表型组档案服务, 为全球研究人员提供更便捷的访问途径。目标是增强 EGA 的服务, 用于共享和存储生物医学研究项目的基因数据和表型数据。这样, 研究团队就能充分利用与他们的研究领域相关的现有测序数据, 从而消除重复工作, 支持先进的研究, 并为小型研究机构提供更低的准入门槛。”

CRG 和欧洲基因组表型组档案系统与运营经理 Ángel Carreño Torres 继续说道: “EGA 包含各种各样不同的数据格式, 比如既有原始的 DNA 核苷酸序列, 也有更细化的格式, 其中包含有关测序如何与人类参照基因组相关联的信息。”

“我们的当务之急是 EGA 档案数据的规模和种类庞大, 且还在迅速增长, 目前, 数据规模已经超过了 2PB。为了有效开展工作, 研究机构需要迅速获取大量数据, 这些数据可能来自不同的小文件, 几个大文件, 或者两者皆有。事实上, 我们最热门的研究之一包含 500 TB 的基因数据。从早期就可以看出, 传统的文件传输协议无法提供所需的性能。”



“与 IBM 业务合作伙伴 TmediaT 合作部署 IBM Aspera, 是推动欧洲基因组表型组档案发展的关键驱动因素。”

--- David Camargo, IT
总监 Centre de
Regulació Genòmica

选择久经验证的解决方案

多年来, EMBL-EBI 一直依赖 IBM Aspera High-Speed File Transfer 解决方案, 来满足内部文件传输需求。在 CRG 加入项目后, 该机构决定进行一次概念验证, 以确定 IBM 解决方案是否依然是实现项目目标的理想平台。

“我们的 EMBL-EBI 同事已经在大范围使用 IBM Aspera, 我们想知道我们能否将 EMBL-EBI 获得的收益扩展到我们的 EGA 工作中来, ” Ángel Carreño Torres 回忆道。

在概念验证阶段, CRG 与 IBM 业务合作伙伴 TmediaT 的专家顾问合作, 对 IBM Aspera 解决方案进行了一次深入测试。在这个阶段, CRG 和 TmediaT 发现、报告并再现了一系列技术问题, 这让该机构能够提前修复这些问题并且降低了项目延期的风险。

“与 TmediaT 团队合作是一次非常愉悦的体验, ”David Camargo 评价道, “在他们的帮助下, 我们确信 IBM Aspera 能够为 EGA 提供所需的性能。此外, TmediaT 召开了一场技术研讨会, 为我们提供了有关 Aspera API 和开发流程的详细信息。投资大型软件组件是一个重大决策, 通过与 TmediaT 合作, 我们确信我们作出了正确的选择。”

高速传输文件

在成功完成测试后, EGA 和 CRG 组成的联合团队在高性能且经过强化的计算环境中部署了 IBM Aspera High-Speed File Transfer。

“DNA 测序研究会迅速生成海量数据。15 年前, 对一个人类基因组进行测序需要花费 30 亿美元和 13 年的辛苦工作, 但是现在, 我们只需投入不到 1000 美元, 就能在几个小时内完成这项工作, ”Ángel Carreño Torres 补充道。

如今, EGA 和 CRG 拥有所需的安全、高性能的平台, 用于管理宝贵的基因数据, 并与世界各地的机构共享这些数据。

“在研究团队收集了来自个人的匿名测序数据且个人签署了同意书授权其数据发布可用于研究用途时, 研究团队就可以通过我们的在线门户将数据上传到 EGA, ”Ángel Carreño Torres 解释道, “然后, 其他研究机构可以申请使用这些数据, 负责监控所收集数据研究的机构会审核他们的请求。”

“一旦请求得到了批准, 我们就会利用为每个请求分配的密钥重新加密数据, 将数据加载到容器内。然后, 请求者能够通过 IBM Aspera 客户端下载容器中的内容, 在安全的存储库中解密数据, 再进行分析。”

解决方案组件

IBM® Aspera® High-Speed
File Transfer

支持前沿研究

借助 IBM Aspera 解决方案，EGA 和 CRG 实现了自己的目标：高速向全球各地的研究团队大规模提供可用的研究数据。

“我们的传输速率取决于访问系统的并发用户的数量；通常，我们每秒会交付 1GB 的数据，当系统需求比平常高时，我们最高每秒能交付 3GB 的数据，”Ángel Carreño Torres 表示，“尽管我们也发现其他的文件传输解决方案也能提供这样出色的性能，但是没有哪一款解决方案能为我们提供像 IBM Aspera 一样的易用性。我们的请求者只需要做一件事：在他们的系统上安装 Aspera 客户端，然后开始下载他们拥有授权的数据。”

Ángel Carreño Torres 继续说道：“对于 IBM Aspera 提供的性能级别，我们再怎么夸大也不为过。”

“有一次，我们发现一个请求者的性能突然下降。于是，我们联系了该请求者，询问我们能否提供任何帮助。结果发现，文件传输太快，占用了他们太多的带宽，使得他们的网络不堪重负，于是，我们必须以较低的速率继续传输文件。我们甚至听说，一些请求者在领略了 IBM Aspera 提供的积极体验后，为保护自己的投资而开始升级网络硬件。”

“未来几年，依赖我们的数据的机构将继续增强他们的计算、网络和存储基础架构。我们坚信，我们的解决方案将帮助他们从这些技术投资中实现最大的价值，因为他们能从 IBM Aspera 的全文件传输速度中受益。”

David Camargo 总结道：“与 TmediaT 合作实施 IBM Aspera 是 EGA 项目的一个重要驱动因素，它能帮助我们实现支持前沿基因组研究的目标。”

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。