

StorNext 与 Scalar 系列

针对 DNA 有关的工作，StorNext 数据管理为基因组测序中心提供强大支持

在数据处理上，很少有工作能比 30 亿物质单元（它们构成了人体 24 种不同的染色体）的测序工作更繁琐。美国贝勒医学院人体基因组测序中心 (HGSC) 在数据的共享、管理和存储方面曾面临巨大难题，但自从部署了昆腾 StorNext 数据管理软件后，一切问题迎刃而解。

传统的技术架构阻碍了研究的进展

美国有三大联邦政府赞助的基因组测序中心，它们致力于研究基因对人类疾病的影响，并不断积累这方面的知识。作为三大中心之一，HGSC 拥有不到 200 名工作人员，包括约 40 名专门分析 DNA 定序数据的研究专家。中心每天会生成大量数据，同时还需要保证数百 TB 的数据可以随时访问以供分析之用，这种情况下，HGSC 陈旧的技术架构已成为开展重要卫生研究工作的绊脚石。

2008 年 8 月，Geraint Morgan 作为信息系统主管加盟该中心，当时的技术已无法满足日益加快的定序数据生成和分析要求。

HGSC 拥有 32 台基因组测序设备，包括 20 台 Applied Biosystems SOLiD 测序机、2 台 Illumina Genome Analyzer 和 10 台 Roche/454 Genome Sequencer。而其中效率最高的设备每天大约可以生成最多 1 TB 的原始数据。

Morgan 面临着这样一种环境：DNA 测序产生的主要数据最初要从本地写入所连接的存储设备（这些设备分属不同的厂商），然后再转至中央存储库。在中央存储库，这些数据由计算节点的群集进一步处理，并通过简单的网络文件系统供研究人员访问。

Morgan 表示：“庞大的数据量让网络基础设施倍受压力，而且也限制了数据的访问能力，这点对于我们这种研究机构来说至关重要。同时，我们需要保证研究人员测序链条操作的末尾不受影响，这样一来如此庞大的数据量也大大加重了管理的负荷。”

为了扩展 HGSC 的存储能力，他们又构建了一个数据中心。但是 Morgan 仍面临着一个难题，就是如何实现对包含服务器、网络和存储技术在内的复杂异构环境实现集中的管理。对于这类政府赞助机构来说，购买新技术面临资金限制，所以他需要对现有的服务器和存储阵列进行扩容，而不是彻底“改头换面”去采用全新的方法。

Morgan 表示“我们需要的解决方案不但要能发挥现有硬件的作用，还要具备轻松的扩展能力以满足今后两年预计 20 PB 测序数据的需要。这套解决方案必须能够在环境发生改变时确保整体管理不受较大影响。”

STORNEXT 在高性能和满足数 PB 环境要求方面的良好声誉是巨大“卖点”

Morgan 最初把目标锁定在多家厂商身上，并通过咨询顾问进行评估，但最终都因为增加工作流程复杂性或无法满足预期发展需要而作罢。



“StorNext 在一款解决方案中集高速数据共享和经济高效的内容保留功能于一身，让我们的研究人员可以根据需要快速地访问数据，并显著降低了原有系统固有的高管理负荷。”

Geraint Morgan
信息系统主管

解决方案概述

- StorNext 文件系统
- StorNext Storage Manager
- Scalar i2000 磁带库系统

主要优势

- 支持同时访问大批量数据而不会影响系统用户
- 通过自动化数据管理提供经济高效的内容创建
- 支持异构环境的集中管理
- 通过整合原有资源保护现有投资
- 针对未来两到三年约 20PB 的预期存储数据增长提供强大的扩展基础

Morgan 表示：“鉴于现有硬件的限制、紧张的预算以及满足数据疯涨的能力，在三个月的寻找中，几乎没有让我感到满意的解决方案。但最终有一款解决方案脱颖而出吸引了我的注意，就是昆腾 StorNext 数据管理软件。”

多家面临相同高性能数据处理需求和大型数 PB 存储要求的公司在使用 StorNext 后都取得了成功，这让 Morgan 欣慰不已。同时他也意识到开源解决方案的低价格和无干扰优势只是表象而已，因为这其中并未考虑到支持合同费用等其他费用。

Morgan 表示：“StorNext 提供了我们所需的扩展能力以及对现有存储硬件的支持，无需额外购买其他硬件，同时它也是一款便于管理的系统。”

HGSC 购买了 StorNext 文件系统和 Storage Manager，实现了多个工作环境下的文件共享以及多个存储层之间的自动化数据移动。

扩展能力强大的解决方案的基础所在

Morgan 表示 StorNext 的部署十分简洁明了。HGSC 目前拥有 2 台元数据控制器和 8 个 StorNext 文件系统 SAN 网关。这些网关通过 4 Gbp FC 网络与存储阵列连接，而计算节点通过经由 SAN 网关的双 4×10 Gbp 以太网链路，借助 StorNext Distributed LAN 客户端访问存储数据。

本地基因组扫描设备收录数据并进行初步预处理后，数据会通过 NFS 复制到中央 StorNext 文件系统。之后，

StorNext Distributed LAN 客户端上运行的多个基因组分析应用程序会执行链条分析，该客户端连接到中央存储位置以并行处理数据。

此外，HGSC 还可使用 StorNext Storage Manager 在不同的磁盘系统与 Quantum Scalar i2000 磁带库之间自动移转数据，从而以较低的成本实现数据保护。以前的基因组项目也可以在 Scalar i2000 上自动归档，从而快速空出主磁盘存储空间供新的测序工作使用。

自从部署 StorNext 后，Morgan 对其带来的优势非常满意，无论是从研究的角度还是 IT 的角度。

他表示：“StorNext 在是一款解决方案中集高速数据共享和经济高效的内容保留功能于一身，让我们的研究人员可以根据需要快捷地访问数据，并显著降低了原有系统固有的高管理负荷。”

展望未来，Morgan 表示 StorNext 将作为扩展能力极强的解决方案的基础，满足未来一年里 2 PB 的预计存储需求以及未来两到三年里最高 20 PB 的预计存储需求。

Morgan 表示：“考虑到基因组研究的性质 — 当日生成的数据或许对将来不具备明显的价值，但有助于带来一些重要发现 — 因此我们会保留所有生成的数据。而这会使存档量不断增加，通过 StorNext 就很好地应对这一问题。同时，疯涨的数据也是我们计划利用 StorNext 发挥重复数据删除功能的优势的原因之一，有了它，我们可以进一步优化存储容量，供存档之用。”

“疯涨的数据也是我们计划利用 StorNext 发挥重复数据删除功能的优势的原因之一，有了它，我们可以进一步优化存储容量，供存档之用。”

Geraint Morgan
信息系统主管

关于人体基因组定序中心

美国贝勒医学院人体基因组测序中心 (HGSC) 成立于 1996 年，是基因组领域的全球权威机构。HGSC 的基本宗旨是通过不断改进基因组技术来推动生物学和基因学的发展。作为由美国国家卫生研究院赞助的全美三大测序中心之一，HGSC 德克萨斯医疗中心为尖端基因组技术在科学与医学领域的应用提供了千载难逢的机会。



Q&U 介绍

我们的目标是帮助您保留全球最重要的数据。Q&U 是指昆腾通过各种途径与您协作，共同解决您面临的数据保护和保留难题，包括分享我们的见解和专业知识，为您提供适当的资源以便您作出明智的选择，同时为您提供全面的解决方案，帮助您应对当前难题并满足未来所需。Q&U 将成为您走向成功的一大“法宝”。

Quantum®

Preserving the World's Most Important Data. Yours.

欲获取联系方式和产品信息，请访问 www.quantum.com/cn，或联系当地办事处。