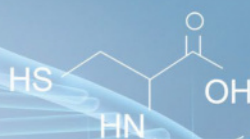
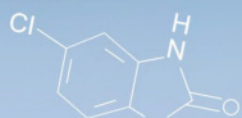


浪潮信息



SOLUTION

DNA基因测序大容量存储

解决方案



方案背景

PROJECT BACKGROUND

我们知道，遗传信息存储在称为基因的DNA序列中。一个人的基因组，它包含的原始数据大小大约是3GB，包含了30亿个碱基，由A、T、C、G四个碱基对组成。在全基因测序过程中，为保证基因数据的完整性，二代测序需要平行测序30次，当最终测序完成，一个人的全基因组的数据会有100G base pair。而生物过程的线索除了DNA信息还有单细胞多组学信息、细胞显微拍照、器官CT、群体层次行为和基因研究等多模态多组学信息。可想而知，采集一个过程中的生物相关的信息将会有非常庞大的数据量。

业务挑战

BUSINESS CHALLENGE

02

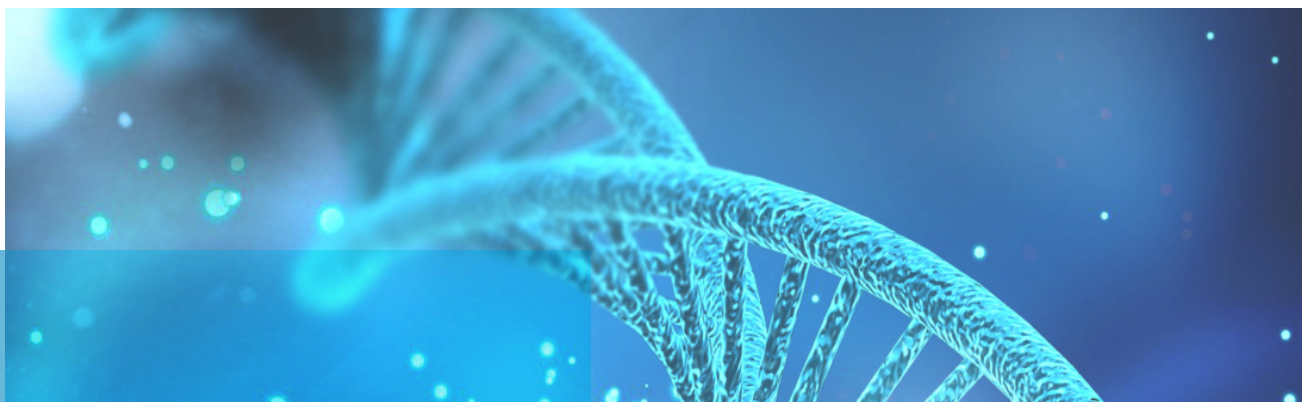
整个基因分析研究分为基因测序、基因数据分析和基因数据归档三个阶段。其一，基因测序的整个过程是不允许被中断的，这就要求支撑测序应用的存储系统具备极致的稳定性和可靠性；另外，基因测序25微秒采集一幅高清晰图像，对性能要求极高；其二，基因数据分析工作中，对存储系统的性能、小文件处理能力提出很高要求；其三，在数据归档阶段，需要将海量基因数据长期、完整、安全地保存起来，存储系统的采购成本、能耗是一个很大的挑战。

解决方案

SOLUTION

03

浪潮信息凭借多年在生命科学基因领域积累的经验，通过Scale-Out横向扩展、IB低延迟网络、私有客户端技术、小文件聚合及使用傲腾盘等异构介质构建冷热数据分层等，满足生物基因应用中对海量数据存储场景高可靠性、高性能、低延迟和低能耗的存储需求。

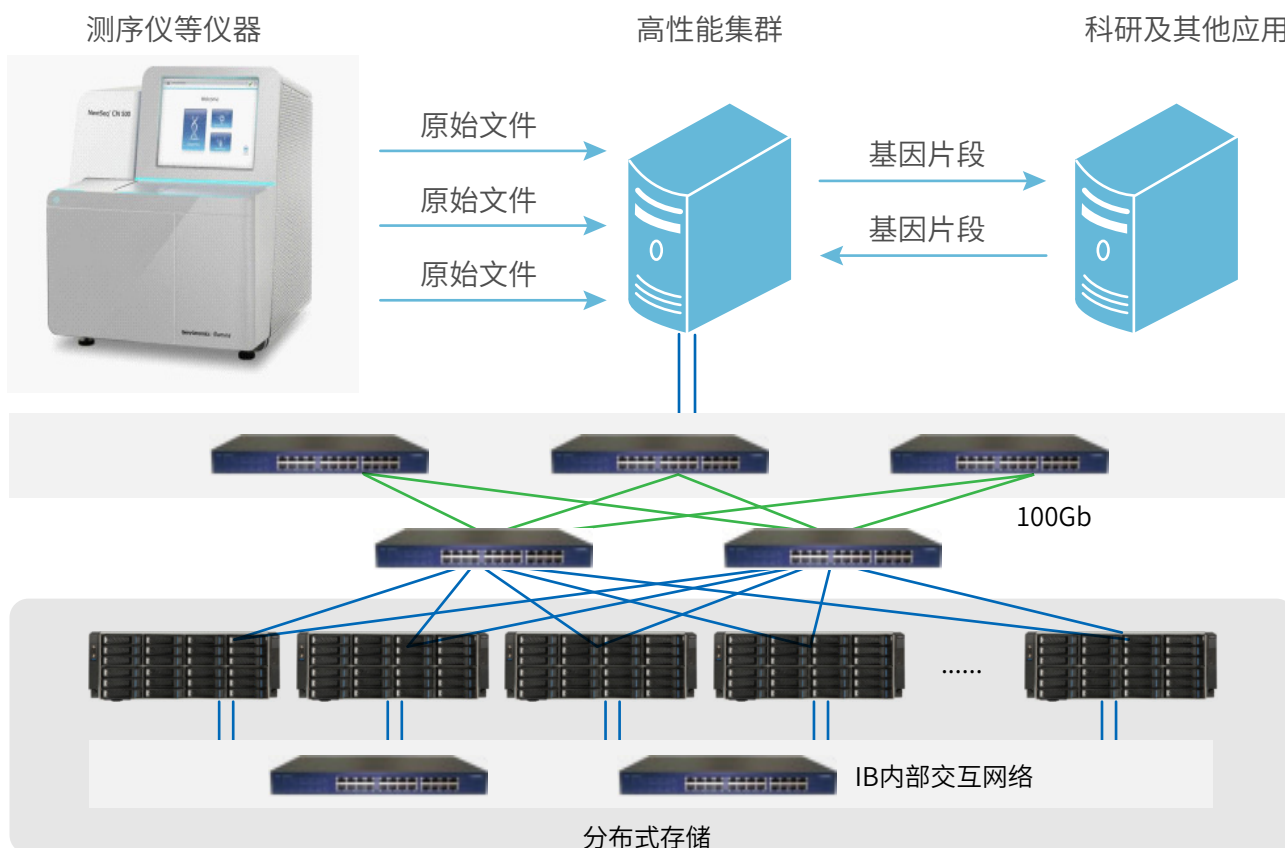


方案亮点

SOLUTION HIGHLIGHTS

04

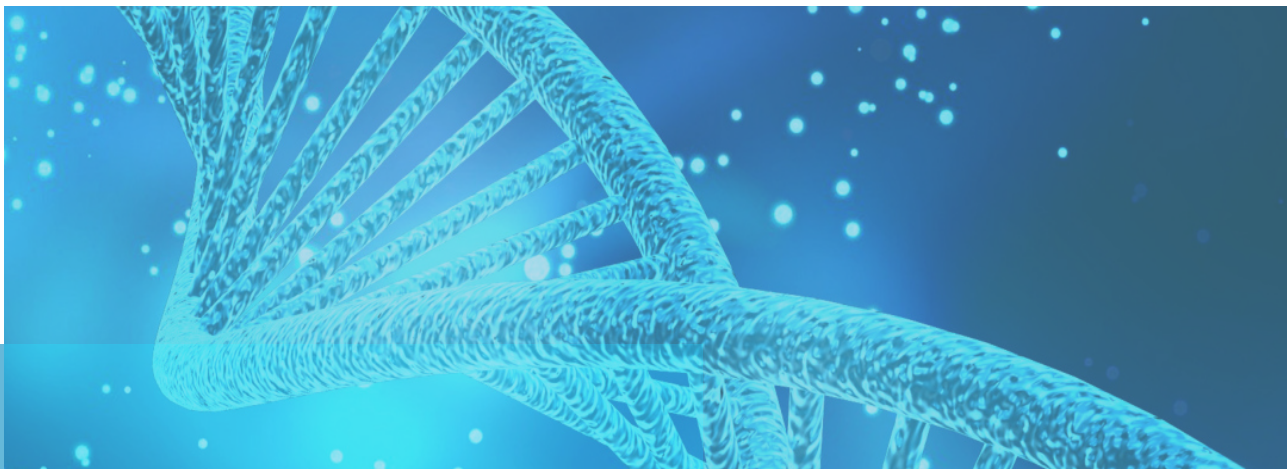
- 通过对架构的优化设计，在文件场景下单一目录可支持1000万文件，文件系统支持千亿级文件，单一命名空间支持20PB；支持5120节点的扩展，单集群可提供EB级容量空间。
- 节点并行处理数据 IO、保障系统整体负载均衡，数据分散存储，避免单节点故障带来的风险，节点在线横向扩展，满足持续增长的容量和性能需求。支持丰富的主机连接接口，支持 10/25/40Gb iSCSI、56/100 InfiniBand 主机连接，无缝接入用户现有环境，满足客户对高带宽及高性能的差异化需求，支持后端 10GE 组网下的 RDMA 模式传输，支持 TOE 卸载，提高系统性能。
- 通过节点级和硬盘级均衡，磁盘间容量误差率不超5%，节点间误差不超1%，可用容量利用率高达95%。同时高密度存储节点满足客户存储空间和能耗要求。



客户收益

CUSTOMER BENEFITS

05



- 生物基因的数据具有极高的价值，分布式存储通过业界领先的冗余设计，为用户提供强大的数据保护能力和系统可用性，数据可用性高达99.9999%。
- 分布式存储通过小文件聚合、对象聚合、全局缓存、多通道、硬件卡加速等技术，为客户提供创新的性能加速方案，全面提升存储系统性能。
- 基因研究产生的数据正以每12-18个月10倍的速度快速增长，分布式存储通过卓越的系统架构，在为用户提供海量的存储空间的同时，通过大比例纠删的功能，大幅提升容量利用率，为客户提供更低的TCO。

浪潮信息电子信息产业股份有限公司 www.ieisystem.com

浪潮信息技术支持与服务热线 400-860-0011

购买咨询热线 400-860-6708 / 800-860-6708

文中有关产品图片及文字仅供参考，详细产品规格及价格，请向浪潮信息当地经销商查询。版权声明©2023浪潮信息，保留一切权利。BJ 06/2023



浪潮信息服务号