

GENOME DATA MANAGEMENT WITH NETAPP



Accelerate sharing and insights with an intelligent data infrastructure.

Turning genomic data overload into insights

The rapid growth of genomic datasets is changing our understanding of human health and disease while enabling incredible breakthroughs in drug discovery, diagnostics, and precision medicine. Individual patients, and even their individual diseases, can now be sequenced to provide insight into personalized treatment plans and targeted therapies. Next-generation sequencing coupled with the power of GPU computing is generating the data and the compute resources to create more-promising candidate medicines faster and more economically than with conventional methods.

The true value of genomic profiling comes from not just one sample but the aggregate of billions, if not trillions, of samples. To accelerate discovery, organizations like yours must retain and manage massive quantities of data to be analyzed and shared around the world.

However, the sheer volume, complexity, and sensitivity of genomic data create significant management challenges for IT organizations throughout the lifecycle of the data—from curation, analysis, and retention to data privacy, protection, and compliance.

A growing opportunity

40EB

Around 220 million genomes (40EB of data) are produced annually.

8 hours

Genomic findings can be returned to critical-care patients in less than 8 hours.

200GB

A single human genome sequence takes up 200GB of storage space.

Empower genomics with intelligent data infrastructure

NetApp can help you accelerate discovery, improve collaboration, and reduce costs throughout the genomic data lifecycle with an intelligent data infrastructure that enables secure, silo-free data management on premises and in the cloud. Together with our partner offerings, NetApp® solutions can help speed genomic research breakthroughs and reduce time to market. You get the control that you need to understand, to manage, and to package your genomic data in an AI- and machine learning (ML)-friendly fashion so that you can more rapidly extract valuable genomic insights.

The NetApp portfolio provides industry-leading intelligent data infrastructure—unified data storage integrated with intelligent data services that deliver security, governance, observability, optimization, and AI performance. By using a NetApp enterprise-grade storage service that's natively embedded in the world's largest clouds, you can seamlessly manage and transfer your organization's genomic data across diverse multicloud environments. NetApp data services deliver active data management through superior data security, protection, governance, and sustainability.

Accelerate discoveries with scalable, high-performance storage solutions

NetApp all-flash storage solutions provide the speed and scalability that you need to handle massive sequencing datasets, enabling rapid data processing and analysis. These solutions enable faster data ingestion, data analysis, and visualization of complex genomic datasets, reducing analysis times from days to hours. The ability of NetApp ONTAP® software to clone datasets by using NetApp FlexClone® technology supports iterative research while preserving original data integrity. A disaggregated storage architecture enables you to scale compute and storage resources separately, and to burst to the cloud if necessary.

Enable global collaboration with unified data management

The more data you have, the more difficult it is to move it around from one analytic platform to another. As genomic data grows in size and volume, it becomes harder to enable seamless data movement and to facilitate the collaboration that is so important for driving scientific breakthroughs.

NetApp is the only storage vendor that offers a unified data storage platform that spans block, file, and object storage and the world's largest clouds. You can easily and efficiently move your data to where you need it, without sacrificing data protection or security. NetApp FlexClone technology enables instantaneous, space-efficient clones of data, so your scientists can collaborate without contaminating the original dataset. NetApp BlueXP™ classification uses AI to help you confirm that no personally identifiable information (PII) is included in your training datasets for AI.

KEY BENEFITS

Accelerate genomic discoveries

- Reduce analysis times from days to hours.
- Rapidly clone data to support iterative research while preserving data integrity.
- Support AI workloads for faster, more accurate patient diagnosis and research results.

Enable global collaboration

- Securely use the public cloud to enable collaboration, with unified data storage that spans block, file, object, and cloud.

Simplify cloud integration

- Maintain consistent data management and control across on-premises and multicloud environments.

Achieve cyber resilience across the data lifecycle

- Proactively detect potential cyberthreats.
- Quickly recover data and applications.

Control costs

- Reduce your storage footprint with thin provisioning, deduplication, and compression.
- Automatically tier cold data from file to object storage on premises or in the cloud.

Simplify cloud integration for genomic workflows

Today's genomic data lifecycle relies on a hybrid cloud model to balance scalability, security, and compliance across its key phases: data generation, processing, sharing, analytics, and archiving. Data may be kept on premises for regulatory compliance and for frequently accessed datasets, but more and more genomic data is heading to the cloud for processing, sharing, and archiving.

NetApp provides seamless cloud integration through partnerships with leading cloud providers, such as AWS, Azure, and Google Cloud. NetApp cloud-native solutions enable your organization to benefit from the elasticity of cloud resources while maintaining centralized control and taking advantage of enterprise storage efficiency and security features.

Your genomic researchers can tap into highly scalable data lakes on premises and use NetApp Cloud Volumes ONTAP® software to present fully curated datasets in the cloud. They can also sequence directly to the cloud, where the data can be curated and retained with all the same security and retention policies as for on-premises storage—in a more cost-effective manner.

Achieve cyber resilience across the data lifecycle

With the most secure storage on the planet, NetApp protects genomic data from sequencing through archiving. NetApp data security solutions protect you from data loss and downtime by proactively detecting potential threats and by quickly recovering your data and applications.

ONTAP enables efficient NetApp Snapshot™ copies, fast cloning, and high-performance random write operations. Erasure coding, encryption at rest and in flight, and immutable snapshots provide a multilayered security and resilience framework that protects you from malware and ransomware. This robust data security framework supports compliance with regulatory standards, such as the Health Insurance Portability and Accountability Act (HIPAA), General Data Protection Regulation (GDPR), and California Consumer Privacy Act (CCPA). Your organization can maintain trust and avoid costly compliance violations.

Control costs with superior storage efficiency technologies

With the trend toward longer genomic data retention periods pushing the size of data volumes higher, organizations like yours need storage solutions to help rein in the cost of exponential data growth. According to the U.S. National Institutes of Health (NIH), genomic data is growing at a rate of 2EB to 40EB annually, driven by advances in sequencing technologies and widespread adoption in clinical settings.

NetApp data efficiency technologies, including thin

provisioning, deduplication, and compression, reduce your storage footprint and the associated costs. Object storage technologies, such as the NetApp StorageGRID® solution, support massive data lakes with Simple Storage Service (S3) compatibility. NetApp FabricPool technology enables automatic tiering of your cold data from file storage to object storage on premises or in the cloud, giving you cost-effective long-term data retention and global accessibility.

NetApp Keystone® Flex Subscription also offers a flexible, pay-as-you-grow consumption model, so your genomic organization can scale resources in line with your needs, without up-front capital investments.

Partnerships with genomic technology leaders

- **PetaGene.** PetaGene's PetaSuite software integrates seamlessly with NetApp technology. PetaSuite shrinks large genomic files to make them more portable, providing immediate data-flow benefits and up to 10 times higher data efficiency.
- **Parabricks.** With NetApp technology and Parabricks from NetApp partner NVIDIA, you can deploy a fully integrated, end-to-end AI solution that's tuned for genomic workloads.
- **NVIDIA.** With all-flash storage from NetApp and NVIDIA, you can significantly boost speed and accuracy for AI-powered precision calculations. NetApp GPU-accelerated genome sequencing is significantly faster compared with CPU-only solutions, maximizing data throughput and minimizing turnaround time.

NETAPP SOLUTIONS IN ACTION



CHILDREN'S CANCER INSTITUTE

For more than 6 years, the Children's Cancer Institute in Australia has worked extensively with NetApp data infrastructure solutions to securely manage over 1TB of research data per patient. The partnership enables the institute to accelerate vital discoveries that contribute to new treatments and, ultimately, help find a cure for childhood cancer.

"By partnering with NetApp, we can better manage and analyze the vast amount of genetic data generated through our research, facilitating collaborations and accelerating discoveries. This partnership brings us one step closer to finding a cure for childhood cancer. As we say, 'It's not if. It's when.'"

*—Anne Johnston, Director of Marketing and Fundraising,
Children's Cancer Institute*

[Learn more](#)

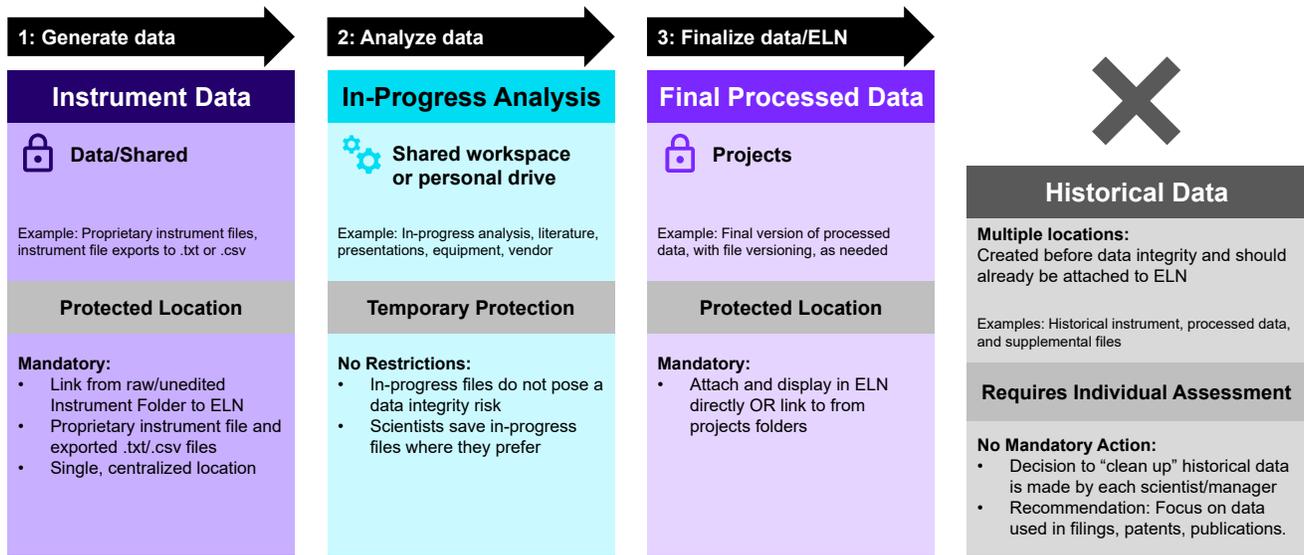


Figure 1) Real-world example of data storage from a major drug manufacturer.

Discover success with NetApp

NetApp has a proven track record of serving the healthcare and life sciences sector with powerful, reliable, and secure solutions that include built-in versioning and traceability to help research and development teams meet compliance regulations. Only NetApp can help you build an intelligent data infrastructure that satisfies the needs of complex, high-performance genomic data pipelines across your entire hybrid multicloud environment.

Learn more about [NetApp solutions for healthcare](#).



[Contact Us](#)

About NetApp

NetApp is the intelligent data infrastructure company, combining unified data storage, integrated data services, and CloudOps solutions to turn a world of disruption into opportunity for every customer. NetApp creates silo-free infrastructure, harnessing observability and AI to enable the industry's best data management. As the only enterprise-grade storage service natively embedded in the world's biggest clouds, our data storage delivers seamless flexibility. In addition, our data services create a data advantage through superior cyber resilience, governance, and application agility. Our CloudOps solutions provide continuous optimization of performance and efficiency through observability and AI. No matter the data type, workload, or environment, with NetApp you can transform your data infrastructure to realize your business possibilities. www.netapp.com

