

Genomics Institute Puts the Right Data into the Hands of Researchers with InfiniteIO

CUSTOMER SUCCESS

InfiniteIO hybrid cloud tiering and metadata acceleration solutions make data accessible at lowest cost and with consistent performance

Customer Spotlight

Company Large Bioinformatics Institute
Industry Life Sciences
Employees 500+

Business Needs

- Make data appear local and accessible
- Identify and move cold data from Tier 1 NAS without interrupting workflows
- Control storage costs at 90% capacity
- Manage 70% Y/Y data growth

Solutions

- InfiniteIO Hybrid Cloud Tiering
- InfiniteIO Metadata Accelerator
- NetApp NAS
- On-premises cloud object storage

Business Results

- Reclaimed 75% of primary NAS by identifying and migrating 2.5 PB of cold data
- Hybrid cloud tiering saves \$1.1 M in primary storage costs for first year & \$7.4 M over 3 yrs
- Offloaded 86% of storage NFS operations from Tier 1 NAS to improve overall performance at scale
- System pricing avoids unnecessary capacity-based management fees
- Data is always available to researchers regardless of physical location



“By intelligently analyzing and managing the institute’s metadata, InfiniteIO identified 75% or 2.5 PB of the file data, which had not been accessed in over six months, to be automatically archived in cloud storage. The hybrid cloud powered by InfiniteIO is saving an estimated \$7.4 million over three years.”

Driving Worldwide Research at Massive Scale

The global scientific community is making major biomedical discoveries and clinical breakthroughs thanks to free and easy access to Big Data by a large European genomics research institute. The institute offers one of the most comprehensive repositories of freely available molecular data resources, training and services to the scientific community; its mission: to drive global genomics research on a massive scale. On average, the institute responds to more than 38 million requests a day over its website, which receives more than 3.3 million unique visitors a month. In 2016, researchers downloaded more than 8.7 petabytes of data.

To make this level of resources available to both the scientific and business community, the institute stores more than 15 years of research data that is accessible at all times on their high-performance network-attached storage (NAS) solution. However, some “cold data” can go unrequested for months or years, while “hot data” might be pulled quickly on a daily basis.

This combination of hot and cold data is difficult and expensive to manage, especially with most of the storage growth coming from data that will likely never be accessed. The institution needed a way to identify and move cold data from expensive tier-one storage without disrupting or changing research workflows. It also needed to improve the way it manages metadata—information about when files were created, accessed and altered—to support rapid data requests and responses.

Optimizing Performance at Lowest Cost

The global organization wanted the cost and scale benefits of a public cloud yet retain the predictable performance and accessibility of on-premises infrastructure. This effort was complicated by the sheer volume of data created and requested daily. The institute’s data is growing 70 percent year-over-year with more than 3 petabytes of data on tier-one storage, which was running consistently at 90 percent of capacity. Much of this data was believed to be inactive but institute officials had no visibility to confirm it.

Even if inactive data could be identified, traditional archive and data management solutions didn’t work because all genomics research data must be kept online and transparently accessible to the research scientists. Traditional tiering and virtualization solutions altered research workflows and could not keep up with metadata performance requirements. The institute’s existing tiering and virtualization solutions were too complex and disruptive to sustain data services to the international academic and commercial research community.

The institute needed to identify and automatically move inactive data from its NAS infrastructure to cost-efficient object storage, without interrupting services to its users. The goal was to make all tiered files available and appear local to support scientific workflows and ensure data storage scales to keep up with growth patterns and performance requirements. The institute needed to deliver consistent storage performance to scientists regardless of where the data is required while reducing infrastructure complexity. After exploring several options, the IT team chose an InfinetIO hybrid cloud tiering cluster along with an object storage system to deliver

a complete private cloud solution that optimizes performance and cost. The institute deployed InfinetIO for transparent discovery, migration and retrieval of inactive data without making changes to its existing IT operations. The solution integrated with cloud storage to reclaim tier-one NAS capacity, minimize storage footprint and spend on cold data, and provide easy access to all cloud-migrated data.

Hybrid Cloud Storage Powered by InfinetIO

- 38 million unique requests per day
- 3.3 million unique visitors per month
- 8.7 PB of downloads by researchers
- 3.5 PB of data generated overall, 2.5 PB of inactive data
- 12 PB inactive data projected in 3 years
- Estimated \$1.1 M savings up-front, \$7.4 M over 3 years

Delivering Always-on Data Services to Global Researchers

The InfinetIO solution optimizes private cloud storage for the institute with greater performance, efficiency and cost savings with zero disruptions to any research scientist or staff. On average, about 86 percent of the organization’s storage requests involve metadata that InfinetIO offloads from its NAS to accelerate responses.

By intelligently analyzing and managing the institute’s metadata, InfinetIO identified 75 percent, or 2.5 petabytes, of the file data that had not been accessed in over six months to automatically archive on object storage. The complete solution is saving the genomics institute \$7.4 million over three years.

Importantly, all data remains securely online and directly accessible to research scientists through their normal workflows, regardless of data location. InfinetIO automatically discovers inactive data and seamlessly moves it to object storage for a durable, active archive that stays ahead of capacity requirements. With the InfinetIO solution, the institute is prepared for future data growth as more genomic Big Data is collected from research. The institute is analyzing and optimizing data to ensure it is assigned to the most logical and cost-efficient areas of the private cloud. Scientists never know how or where the data is stored. They only know it’s quickly available when they need it the most.

InfinetIO makes hybrid cloud storage simple and fast, ensuring information flows to people and applications that need it most, without disruption. Based in Austin, Texas, InfinetIO provides real-time insights to analyze, optimize and scale capacity and performance throughout the hybrid cloud for global enterprises, research organizations and media companies. InfinetIO offers customers and partners the fastest way to implement a hybrid cloud strategy and maximize the value of cloud storage. Learn more at www.infinite.io, [@infiniteio](https://twitter.com/infiniteio) and [LinkedIn](https://www.linkedin.com/company/infiniteio).

©2019 InfinetIO, Inc. All rights reserved. InfinetIO is a trademark of InfinetIO, Inc., registered in the United States and other countries. All other brand names mentioned herein are for identification purposes only and may be the trademarks of their respective holder(s).