Quantum®

Eureka! Cracking the 'Omics Code with StorNext at SIB Swiss Institute of Bioinformatics

At the forefront of genomic and proteomic research, the SIB Swiss Institute of Bioinformatics generates 30TB of data per week. As 'omics and genomics in particular move toward the point of patient care, StorNext scale-out storage gives SIB a proactive strategy for harnessing genomics data for decades.

SIB Swiss Institute of Bioinformatics

FEATURED PRODUCTS

StorNext® 5

"If you provide researchers with the right set of tools, they push the envelope. StorNext® tiered storage helps us take data in fast, quickly move it to archive, and keep it ready so bioinformaticians can continue their work."

Professor Ioannis Xenarios
Director Vital-IT Group, SIB Swiss Institute of Bioinformatics

Vital-IT
High Performance
Computing Center

SOLUTION OVERVIEW

- Quantum StorNext Scale-out Storage
- StorNext AEL6000 Tape Archive
- StorNext AEL500 Tape Archive

KEY BENEFITS

- **Provides high-performance ingest** to accelerate genomics workflows
- **Scales performance and capacity easily** to keep up with 30TB/week of data growth
- **Balances performance and cost** with primary and archive storage tiers that keep research data ready for re-use
- **Automatically protects valuable genomics** data with multiple archive copies
- **Ensures best-in-class data integrity** with extended data life management feature
- **Supports next-gen object storage** and cloud storage tiers from Quantum, making it easy to scale

The independent, non-profit SIB Swiss Institute of Bioinformatics is at the forefront of the next great revolution in life sciences—applying computational methodologies and large-scale data analysis to genomics, proteomics and other bioinformatic sciences. Founded in 1998, SIB includes some 60 bioinformatics research and service groups and some 700 scientists from the major Swiss schools of higher education and research institutes. SIB's dedication to advancing genomics education and research is one of the reasons that Switzerland has the highest concentration of bioinformaticians of any nation on the globe.

SIB's work is increasingly focused on applied genomics. Personalized medicine, population genetics, the biology behind flavor perception, methods for increasing crop yield—all can improve quality of life for us all.

"SIB recently worked on an algorithm for a prenatal diagnostic test for conditions like Down syndrome," explains Professor Ioannis Xenarios, Director at the Vital-IT Group, part of SIB tasked with designing and supporting the innovative computing infrastructure that members rely on for research. "With a simple blood draw from the mother at 11 weeks, we can sequence the genetic material of the fetus in utero. It's less invasive—and much less risky than traditional amniocentesis. And it shows how genomics is becoming more relevant in our everyday lives."

MORE THAN 30TB PER WEEK CREATES UNIQUE DATA MANAGEMENT CHALLENGES

As the applications for genomics increase and the costs and cycle times for sequencing decrease, organizations are running more genomic sequences—and generating massive amounts of valuable data.

"As we transition toward applied bioinformatics, we need to plan for the long-term preservation and maintenance of data—both in terms of scalable capacity as well as associated costs like headcount, energy, and cooling," says Xenarios.

SIB operates six different sequencing centers and supports projects from about 300 active research teams. Sequencing runs take days, and the team typically processes five separate projects each week. The raw data is run through different analytic applications in a pipelined workflow that result in summary tables and graphs suitable for reports and publications. With sequencing generating up to 30TB of data a week, data adds up very quickly.

"Over the last few years, sequencing has become much faster," explains Roberto Fabbretti, Senior Scientist and IT Manager at Vital-IT. "That means we are doing more projects than ever and our data is growing very rapidly."

VALUABLE DATA, LONG LIFESPAN OF RESEARCH DEMANDS DATA STEWARDSHIP

The nature of the cutting-edge research that SIB supports means that Xenarios and his team are, in effect, data stewards throughout the long lifespans typical of genomics research.

"For research into areas like cancer and immunotherapy, we capture large amounts of sequenced data for each patient," says Xenarios. "If that person comes back on a week-to-week or month-to-month basis, all the data from previous tests needs to be made quickly and accurately available to researchers in a short amount of time. To scale our bioinformatics efforts to support tens of thousands of patients, we need to look for cost-effective ways to preserve genomic data for 20, 30, or 40 years of time—effectively creating a view of a patient from before birth to death."

HIGH-PERFORMANCE STORAGE FOR GENOMICS AT PETASCALE

Vital-IT today supports its research infrastructure with StorNext scale-out storage from Quantum. Researchers get high-speed access to sequencing and analysis data through four separate StorNext systems—nearly 1PB of primary storage and 4PB of economic tape archives. StorNext supports high-performance processing for genomics data using IP over

Infiniband. SIB's tiered approach keeps active data on primary storage for complex analysis and automatically moves data into the long-term archive as it ages. Over 600 users access the sequenced genomic data locally by tapping into the network in one of the SIB-affiliated data centers, as well as remotely through a CIFS interface.

"Eight years ago when we started looking into solutions, Quantum StorNext was the only solution that really added value to our way of working," says Xenarios. "We didn't need to change our infrastructure, and a single full-time employee can manage the storage infrastructure. That's a huge benefit in making sure our budget stays focused on supporting our researchers."

SELF-SERVICE ACCESS KEEPS GENOMICS DATA READY FOR RESEARCH

"The data that our researchers capture and analyze provides important answers today, but it also has the potential to be useful months or years later when new analytic applications can extract information from the same raw sequences," Fabbretti says. "StorNext allows us to provide cost-effective long-term archiving for all our projects, regardless of how long a project is planned to last."

Once a research project has passed the active processing stage, the SIB workflow automatically moves files from primary disk to Quantum StorNext AEL tape archives in a process invisible to researchers. Once a file has been tiered to the archive, it still appears where the researchers expect it to be in the file system, as if it were still on disk. And IT doesn't have to wade through requests to recover archived data. Self-service access means that researchers can easily access files that have been archived, without needing to file an IT support ticket.

"If you provide researchers with the right set of tools, they push the envelope," says Xenarios. "They sequence 1,000 people, and you get a massive load of 800TB in a short few months. StorNext tiered storage helps us take data in

"With life sciences data, it's incredibly hard to know where you're going to be in five years, especially as physicians become the next generation of data scientists. StorNext allows you to keep that information for a very long time in a cost-efficient way."

Professor Ioannis Xenarios,
Director Vital-IT Group, SIB Swiss Institute of Bioinformatics

ABOUT THE SIB SWISS INSTITUTE OF BIOINFORMATICS AND THE VITAL-IT GROUP

Including some 60 bioinformatics research and service groups and some 700 scientists from the major Swiss schools of higher education and research institutes, the SIB Swiss Institute of Bioinformatics is an academic, non-profit foundation established in 1998. SIB coordinates research and education in bioinformatics throughout Switzerland and provides bioinformatics services to the national and international life sciences community. SIB's Vital-IT Group provides computational resources, storage infrastructure, development support, and bioinformatics expertise to help SIB's scientific community conduct research, and it supports clinical practices for associated medical facilities.
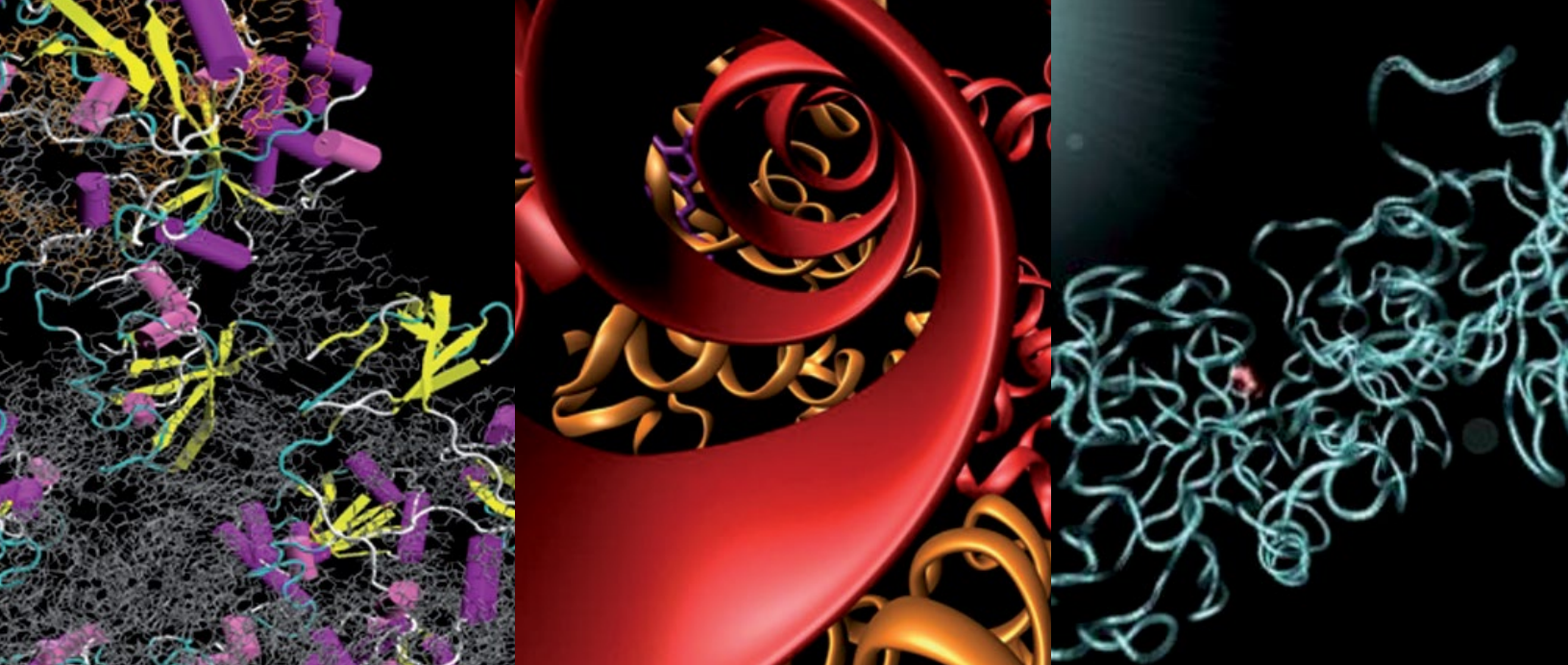
fast, quickly move it to archive, and keep it ready so bioinformaticians can continue their work."

AUTOMATED PROTECTION FOR SOME OF THE MOST VALUABLE DATA SETS ON EARTH

"StorNext not only helps us make sure we capture data fast—it also makes archiving an automated, cost-effective process to help us fulfill our role as a data steward," says Fabbretti. "We always make two copies of the files on tape, keeping one available in the archive and the other vaulted to provide an additional layer of protection against any kind of hardware failure or damage to a site."

The protection also extends to data in the Quantum archive, which provides best-in-class management, monitoring, data integrity and data security capabilities. Extended Data Life Management (EDLM), a key feature of the Quantum tape archives, periodically loads tapes into special drives and checks the media and

the data stored on them. If suspect media is detected, the data is automatically written to new media to maintain the integrity of the information.

"We are dealing with some of the most valuable data sets on earth," Fabbretti explains. "StorNext gives us a multi-petabyte archive capability, long-term data protection, and the ability to easily roll back file versions—it's a critically important part of that strategy."

SCALABLE PERFORMANCE AND CAPACITY KEEP SIB READY FOR WHAT COMES NEXT

While the fields of genomics and proteomics are changing quickly, the rapid increase in data is a constant. With scalable performance and capacity, StorNext keeps SIB ready to support whatever innovations come next.

"StorNext has supported our growth for over six years. We know we can easily add more disk and capacity when we need it. In

fact, we've gone beyond genomics to store and protect general medical research data sets. It is important to us that StorNext can easily include additional tiers like cloud or object storage in our storage workflow when it is time to expand."

SIB's long experience in supporting genomics research has put it into a thought leadership role for the institutional partners it works with as they get their own life sciences IT programs up and running.

"People are coming to us for advice on what to choose in terms of technologies," says Xenarios. "With life sciences data, it's incredibly hard to know where you're going to be in five years, especially as physicians become the next generation of data scientists. Data tends to accumulate fast and you can't throw it away. StorNext allows you to keep that information for a very long time in a cost-efficient way."