

StorNext

## The Swiss Institute of Bioinformatics Reduces Cost of Multi-Petabyte Storage by 50% with Quantum StorNext Software

When The SIB Swiss Institute of Bioinformatics was faced with spiralling data growth arising from next generation sequencing, it deployed a hierarchical storage management (HSM) solution centered on Quantum StorNext data management software and HP hardware. This provided high performance file sharing and data protection and reduced SIB's total cost of storage by 50%.

### HUGE DATA VOLUMES PUT PRESSURE ON STORAGE INFRASTRUCTURE

SIB is a federation of bioinformatics research and services groups from leading Swiss universities and the Swiss Federal Institutes of Technology. The Vital-IT Center is the heartbeat of SIB, acting as a high-performance computing (HPC) joint venture between academic and industrial partners. Dedicated to life sciences, the Center supports software development and optimisation along with HPC and data storage for biology and medicine.

Bioinformatics—the process of applying information technology to biological problems—creates an ever-increasing volume of data. An example is the DNA sequence analysis needed to support the worldwide Human Genome Project.

Next generation, ultra-high throughput sequencing allows genomes to be sequenced in an unprecedented manner. A single experiment produces up to 743,000 files per run, with each run sized at an average of 2 TB and performed every 3.5 days. The performance requirements relating to these scans are equally large: a 600 MB/sec aggregate read and 400 MB/sec aggregate write are mandatory to allow data analysis.

Over time, the data per run had grown significantly at SIB. In 2007, 1 TB of raw and processed data was produced during each week — by 2009 this had risen to 7TB per week and was continuing to grow. This put tremendous pressure on the Vital-IT Center to store, protect, and manage the data as existing storage capacity and budget were no longer sufficient.

"We needed to create a storage infrastructure capable of scaling to multiple petabytes and managing hundreds of millions of files," says Roberto Fabbretti, IT Manager at the Vital-IT Center. "We wanted to eliminate the need for

a separate backup and provide a comprehensive disaster recovery solution. All of this needed to be wrapped within a cost-effective storage environment."

The Center considered traditional storage area network (SAN) and network attached storage (NAS) options. A SAN would have enabled SIB to store data centrally and access that data quickly; however the costs of deploying a fiber channel-connected infrastructure were found to be prohibitive. A NAS solution was also eliminated from consideration because of the costs involved, as well as the limited performance and scalability it would have provided.

In addition, both options would still not have provided the required data protection - keeping track of original data for up to 20 years is particularly important for the pharmaceutical and biotechnology industries.

Rejecting the two options, Vital-IT discussed the situation with partners and decided to choose an HSM solution that would move data automatically between hard disk arrays and tape storage and would provide flexible, high speed access to many users. StorNext emerged as the clear choice.

### STORNEXT MEETS HPC FILE SYSTEM AND HSM REQUIREMENTS

"We examined a number of solutions," explains Fabbretti. "StorNext met our combined HPC file system, HSM, and data protection requirements. This transport- and hardware-independent solution offered scalable performance and easy resizing of volumes, as well as water-tight disaster recovery protection for the 400 TB of data we were managing."

The Vital-IT Center deployed a StorNext solution comprising StorNext File System and StorNext Storage Manager. StorNext is currently integrated within an HP



High Performance Computing Center

"When it comes to efficient file sharing, transparent tiered storage, and cost-effectiveness, StorNext has fulfilled all our requirements."

#### Roberto Fabbretti

IT Manager, the Vital-IT Center, part of The Swiss Institute of Bioinformatics

### SOLUTION OVERVIEW

- Quantum StorNext File System
- Quantum StorNext Storage Manager
- HP BL480 metadata servers
- Three HP BL680 SAN gateways
- HP disk cache with 160 TB of extensible storage
- HP tape library scalable to 570 TB

### KEY BENEFITS

- Reduced total cost of storage by 50%
- Improved sequence tag productivity by 20%
- Provided operational cost savings, including cooling and power
- Enabled high-speed data sharing with cost-effective content retention
- Consolidated resources and enabled workflow operations to run faster
- Moved data between storage tiers transparently for simplicity, scalability, and economy
- Eliminated vendor lock-in through StorNext platform independence
- Ensured all files were easily accessible to all hosts

disk and tape library environment which includes two HP BL480 StorNext metadata servers, three HP BL680 SAN gateways, an HP disk cache with 160 TB of extensible storage and an HP tape library that can scale to 570 TB.

## "STORNEXT HAS TRANSFORMED OUR DATA STORAGE INFRASTRUCTURE"

Since deploying StorNext, Fabbretti and others at the Center have been very pleased with the results.

"StorNext has transformed our data storage infrastructure," he says. "Our first file system spans 22 million files, and the second file system we recently purchased already has 2.5 million files."

The StorNext File System provides high performance access to the central storage pool for multiple users, and because it is heterogeneous, applications on different operating systems can be used to collect, store and analyse the data, all at the same time.

File system performance is optimised through the use of StorNext "affinities" which organise the writing of data from one physical disk to another.

Access is provided over GbE using industry standard protocols (NFS / CIFS) and also over SIB's InfiniBand network providing high performance and resilience. In the future, SIB is also likely to deploy the StorNext performance protocol, Distributed LAN Client, to provide even higher levels of performance access for key users.

When dealing with large data sets, it isn't cost-effective to hold petabytes on high performance disk, and yet the management and administration costs of moving data to cheaper storage can be prohibitive unless an automated system like StorNext can be utilised. Figure 1 shows the

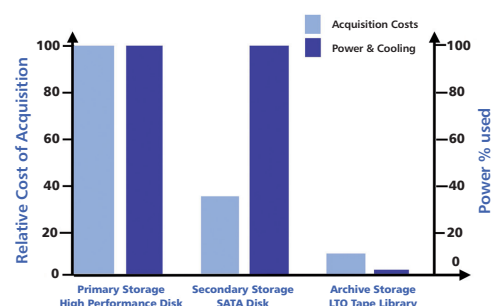


Figure 1

cost of storage in relation to purchasing and ongoing operational costs whilst Figure 2 shows data in relation to its value in terms of frequency of access.

In the case of DNA sequencing, high performance storage is required to collect and process sequence data. However once results have been generated, the data is suitable for storage on more cost-effective tape and secondary disk.

As data is stored in the file system, StorNext Storage Manager also copies it to other tiers of storage based on predefined policies. Over time, unused data is removed from the primary storage, leaving just the data held on the other tiers. File stubs are left in the file system, so users can access data from the same place that it was originally stored.

Data is continuously protected by making copies as it is written to primary disk, thereby eliminating backup window requirements. Also, there is no need for additional backup hardware or software, saving further costs.

In the event of a disaster, SIB doesn't have to restore the complete 400 TB of data before it can be used, which would take months. Instead, the system is brought up in a nearline state with project data available as and when needed.

StorNext has also allowed researchers to keep the raw data that most other sequencing centers have to discard—saving 20% on sequence tags as new algorithms are run on original data.

"By standardising on StorNext, we have reduced our total cost of storage by 50%," Fabbretti reports. "When it comes to efficient file sharing, transparent tiered storage, and cost-effectiveness, StorNext has fulfilled all our requirements."

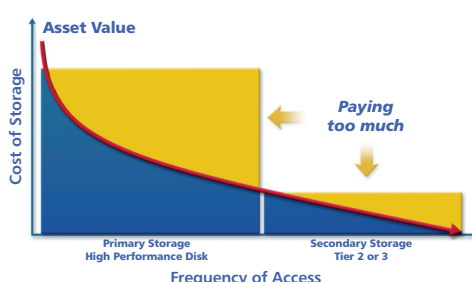


Figure 2

"By standardising on StorNext, the Swiss Institute of Bioinformatics has halved its total cost of storage."

**Roberto Fabbretti**  
IT Manager, the Vital-IT Center, part of  
The Swiss Institute of Bioinformatics



**Swiss Institute of Bioinformatics**

## ABOUT SIB—THE SWISS INSTITUTE OF BIOINFORMATICS

The SIB Swiss Institute of Bioinformatics is an academic not-for-profit foundation federating bioinformatics activities throughout Switzerland. Its two-fold mission is to provide world-class core bioinformatics resources to the national and international life science research community in key fields such as genomics, proteomics and systems biology; as well as to lead and coordinate the field of bioinformatics in Switzerland. It has a long-standing tradition of producing state-of-the-art software for the life science research community, as well as carefully annotated databases.

The SIB includes 29 world-class research and service groups, which gather more close to 400 researchers, in the fields of proteomics, transcriptomics, genomics, systems biology, structural bioinformatics, evolutionary bioinformatics, modelling, imaging, biophysics, and population genetics in Geneva, Lausanne, Berne, Basel and Zurich. SIB expertise is widely appreciated and its infrastructure and bioinformatics resources are used by life science researchers worldwide.



## What is Q&U?

Our goal is to preserve the world's most important data. Yours. Q&U is Quantum's approach of collaborating with you to address your specific data protection and retention challenges. It's about sharing our insights and expertise, giving you the resources to make educated choices, and delivering comprehensive solutions that help you tackle today's challenges while preparing for tomorrow. When Q&U come together, great things happen.

**Quantum**

Preserving the World's Most Important Data. Yours.

To contact your local sales office,  
please visit [www.quantum.com](http://www.quantum.com)