BCM®
Baylor College of Medicine

Baylor College of Medicine

BAYLOR COLLEGE

OF MEDICINE

SECURES CRITICAL

MEDICAL

BIOINFORMATICS

WITH QUALSTAR

XLS LIBRARY

XLS

QUALSTAR®

QUALSTAR

BCM®
Baylor College of Medicine

Baylor College of Medicine

The XLS stores over 1 Petabyte of uncompressed data and is able to transfer almost 7 TB of information per hour, which enables the Center to hit their backup windows, substantially reduce floor space and reduce the amount of tape handling by people and robotics.

THE CENTER

The ongoing war between humans and disease takes place in medical research laboratories where the analysis of enormous quantities of data means the difference between continued suffering and a long-awaited cure.

Rated among the top 10 US medical schools in the country by *US News and World Report* Baylor College of Medicine in Houston, Texas, is one of the main battlefields in this war. It is the only private medical school in the Southwest, and is recognized nationally for excellence in education, research and patient care.

A cornerstone of the College's research foundation is the Human Genome Sequencing Center, established in 1996 during the crucial final phase of the Human Genome Project. It is one of the three sites selected nationwide to complete the final phase of the project.

Finding a single gene amid the vast stretches of DNA that make up the human genome – three billion base-pairs worth – requires analyzing, managing and protecting thousands of terabytes of data. The Human Genome Project (HGP) is devoted to developing new and better tools to make gene hunting faster, cheaper and practical for almost any scientist to accomplish.

Mapping even one gene is a daunting job, and the data generated is staggering. The complexity of the task can be gauged from the fact that the blueprint of a single chromosome contains more than 33 million individual pieces of information.

In addition to human genome sequencing, the Center works with researchers from around the world to sequence the genomes of animals and insects, and bacteria known to cause infection.

Baylor College of Medicine supports the Center's multiple research laboratories with an extensive system of Linux clusters providing both primary processing and data protection services. As research results accelerate, data has increased more than ten-fold.

David Parker, Director of Informatics Systems at the Center, notes: "Preparing DNA generates lots of data. We have three new platforms for sequencing, so we are streaming more than one-and-a-half terabytes per day per machine."

Parker had two major problems to deal with as data came racing in. One was the backup technology itself. "The technology we were using was getting old," he said. "Our capacities were outgrowing LTO 1, and our quarterly backups took two months."

But the technology was not the only challenge. Space was also a consideration. "We've been running out of floor space," says Parker. "We had to leave three feet of clearance on three sides just to open the doors (on the tape backup subsystem they had used)."

Floor space is becoming an increasingly challenging problem at HGSC, and many other facilities. As the volume of data grows, so does the hardware required to store and manage it. Gartner, Inc. has published a forecast stating that, by 2011, more than 70% of US enterprises will face tangible IT disruptions related to energy costs and floor space. Floor space, air conditioning, and power requirements will limit growth in many installations. Costs of power tend to vary by geography and the utility provider. Costs of floor space tend to differ by geography as well, but it is uniformly expensive…and sometimes there is simply no more space to be had.

The HGSC is housed at the Texas Medical Center, a 1000-acre megaplex that holds not only Baylor, but 45 other institutions…government agencies as well as private not-for-profit health institutions. Although the complex seems large, it supports 5.5 million patients per year, offers 6,500 beds and counts over 73,000 employees of the various institutions head-quartered there. The Center is on the 15th floor of their building, and space is at an increasing premium.

With the twin challenges of data growth and finite floor space squeezing the Center, Parker reviewed a variety of backup solutions. Working with systems integrator Teqsys, Inc., he determined that his best value was Qualstar's XLS 820500 Enterprise Library System. As currently configured, the XLS contains 1,340 cartridge slots and a bank of sixteen LTO 4 fibre channel tape drives.

This XLS stores over 1 Petabyte of uncompressed data and is able to transfer almost 7 TB of information per hour, which enables the Center to hit their backup windows, substantially reduce floor space and reduce the amount of tape handling by people and robotics.

With new genetic data streaming in at almost 2 TB per day per sequencing platform, the Center welcomes the advantage of the XLS's density; over 1 PB (native) is stored in just 17 square feet. One of the space saving characteristics of the library is the robotic assembly. The library robot travels an average of just 21 inches horizontally to access tapes. The robotics always remain within the basic library cabinet, eliminating the power consumption, reliability problems and delays that can occur with robots that have to traverse at high speed between cabinets to access remote cartridges.

Parker reflected that installing the library was easy since Qualstar oversaw the hardware installation. The Center uses Veritas NetBackup, and the XLS operates effectively under the newest version of the software.

QUALSTAR

"Performance is much faster in moving tapes," said Parker. "And it gives me far more capacity in the same footprint. Backup capacity has been increased tenfold. And, we haven't had to adjust the robotics either…this was important to us."
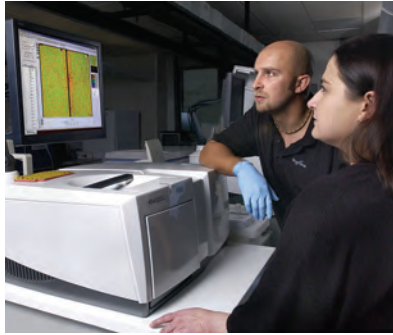
In reviewing options for backup, Parker rejected the notion of using hard disk. Two considerations weighed against the option: "Our computer room is on the 15th floor," he pointed out. "The weight of an array of the size needed would be too great for this floor. Disk drives are heavy, and generate lots of heat. Additionally, our archive has to be available years down the road. Some of this data must never be lost, so we rotate data offsite quarterly. Disk is simply impractical for a number of reasons."

Parker realized, like many, that real world advantages result from using tape as a backup media. The Center rotates their critical data offsite, so portability is a must. In his book, *Disaster Recovery Planning*, Jon William Toigo notes: "Many experts believe that effective off-site storage of critical data is the single most important determinant of successful business recovery following a disaster."

Too many disk-to-disk backup solutions reside in the same SAN as the primary disk storage. One disaster, be it fire, flood, hurricane, power surge or the like, would make both the primary data source and the backup data source irretrievable at the same time. Tape cartridges, on the other hand, are easily dismounted from the tape library and shipped by a variety of carriers to secure off-site facilities. A tape solution is a must, then, for disaster recovery best practices, where portability and cost-efficiency are stated goals.

Scalability is a continuing data center concern with the meteoric rise in the amount of data the average data center must store and retrieve. And scaling is that much harder with the Center's finite floor space. Disk scales by the addition of drives (and controllers, software upgrade or replacement, and floor space); tape scales by the addition of cartridge modules. Does this mean that tape scales to infinite capacity? Probably not, but by contrast to disk the scaling is easy and tape cartridge capacities keep growing. The LTO 4 technology in the XLS deliver a native capacity of 800 GB per cartridge. Innovation in tape technology continues to reduce its cost per GB, maintaining it as the least expensive solution for storing large amounts of data over extended periods of time, contributing to a very positive total cost of ownership.

Scalability is at the heart of the Center's future. As gene sequencing platforms generate data at high speed and ever-higher quantity, data protection will only grow in importance.

"We plan to continue expanding with XLS," says Parker, "until I use up the room that I have." The Center's current XLS 820500 is easily expanded to house more than 6,000 tapes and over 80 tape drives, a five-fold increase in both capacity and data throughput.

The XLS Enterprise Library System is available exclusively through Qualstar's worldwide network of Authorized Resellers. For more information contact Qualstar, one of Qualstar's Resellers or visit:

http://www.qualstar.com/xls

All information and specifications are based on our knowledge of the products at the time of printing. Qualstar reserves the right to change specifications without notice.

Qualstar and the Qualstar logo are registered trademarks of Qualstar Corporation. LTO is a trademark of Hewlett-Packard, IBM and Seagate Technology. All other trademarks are the property of their respective owners.

QUALSTAR CORPORATION

3990-B Heritage Oak Court • Simi Valley, CA 93063 USA • Fax 805-583-7749 • Tel 805-583-7744 • Toll Free 800-468-0680
E-mail: sales@qualstar.com • Website: www.qualstar.com • NASDAQ: QBAK

QUALSTAR® *Simply Reliable*