

WHEN AN INDUSTRY REQUIRES Multiterabyte Data Protection

Ron Levine

The biotech industry, more than most others, depends on its IT systems' stored data for survival. The vital research and test results stored in such systems must be 100% protected. At best, lost or corrupted data can be expensive — costing time, labor, income, and customers. At worst, damaged or lost data may be irreplaceable, threatening business survival. Biotech companies cannot afford to lose data that may have taken years to collect.

According to a recent study by StorNet, Inc., (www.stornet.com), 28% of all system downtime is caused by server disk crashes. That puts the server's critical data at risk and makes it unavailable. Yet the network administrator's backup window continues to shrink. As the amount of data — and the need for backup and protection — grows, faster and more robust backup and recovery techniques are required.

Lynx Therapeutics, Inc. (www.lynxgen.com) develops and applies technologies for discovery of gene expression patterns and genomic variations for the pharmaceutical, biotechnology, and agricultural industries. The base technology, Megacclone, transforms a sample containing millions of DNA molecules into millions of microbeads, each carrying approximately 100,000 copies of one of the DNA molecules from the sample. This is a data-dependent procedure requiring large numbers of huge files, which must be both protected and available.

Lynx is a typical example of a biotech business with crucial (and continually growing) data backup and recovery needs. The data acquired, and its analysis, occupies an enormous amount of online storage that needs to be readily available, reliably stored, and completely recoverable if a disaster such as a fire or earthquake should occur. Fully protecting the company's data requires constant storage consolidation and sophisticated backup solutions.

Lynx uses a backup and data storage management solution (designed by StorNet) that comprises data handling software, RAID storage, and an online library of high-capacity tape cartridges. The developer chose tape media for data backup because it offered an efficient and cost effective storage medium. The company deals with very large file transfers so it requires a high-reliability, large-capacity medium. The media must also be easy to handle and accommodate both online (backup) and offline (archived) storage, and it must be fast enough to accomplish huge data transfers in a reasonable time. Tape is versatile, small, and easily stored, and thus works well for backing up data in an online library and archiving off site.

Lynx's 100 Mb fast ethernet network consists of about 30 servers ranging in size and capability from a Dell PowerEdge 1400, to several multiprocessor Network Engines-based PCs, up to a four-processor SUN E450. The machines have cumulative totals of over 30 GB of RAM and some 3 TB of hard disk storage. They run NT Server, Solaris 2.5.1 and 2.6, and Red Hat LINUX. The Veritas NetBackup software resides on an NT server (for all company desktop backups), on an old SUN Ultra-2 (for the UNIX-LINUX desktop backups), and on the E450 (for the multiterabyte RAID backups).

All production data is stored on a RAID device (for high availability) and backed up onto an STK model L180 automated tape library with two DLT drives and a 180 tape cartridge capacity (35 TB native or 70 TB compressed storage per tape).

The RAID storage unit, an LSI 3702, level 5, with 2 TB (usable) capacity is constantly between 65% and 90% full. "We move data from the RAID to archival tape and place the tape media offsite for safety. About 300 tapes are on-the-shelf at any given time," according to Malcolm Raff, systems and networks administrator at Lynx.

RAID-to-tape incremental backups are performed several times each week and usually involve approximately 100–200 GB of data. Such backups require about six to 12 hours. According to Raff, the company employs a sliding window methodology — it backs up only new data. The backups have a 60-day termination date. The tapes are then recycled and the most recent data is recorded, along with any older data left on the RAID after 60 days. A full network backup of every machine in the company (requiring three days) is performed every month and protects approximately 4 TB of data. When necessary, backups are performed more frequently.

After three years of operation, Lynx Therapeutics is pleased with the high-availability, data protection, and file recovery solution.

Planning for the future is an important part of the company's backup strategy. Lynx has developed a suite of applications (based on Megacclone) that they hope will enhance the pace, scale, and quality of genomics and genetics research programs. They are developing additional applications aimed at addressing the needs of the pharmaceutical, biotechnology, and agricultural industries and are working on a new separation technology to provide high-resolution analysis of complex protein mixtures from cells or tissues of interest. The company also intends to apply its suite of technologies in selected biological areas to develop products internally. As such plans become reality, larger data backups will be required. Its backup solution is scaleable.

To guard their valuable data assets, biotechnology companies must be fully prepared for any potential data disaster and recovery scenario. It's always easier to prevent a data disaster than to recover from one and having extra copies of critical files safely stored on reliable tape backup media is a time-proven data protection method.

Ron Levine is manager of client services and chief technical writer at Coast Writing, Carpinteria, CA, 805.566.9747, www.coastwriting.com. He also teaches computer information systems and multimedia at Santa Barbara City College. ron@coastwriting.com