On Jan. 31, 2007, the total number of trades and quotes on the New York Stock Exchange (NYSE) hit more than 133 million. On the same day but one year later, that number had more than tripled, surging to more than 494 million. Such numbers present a formidable challenge to banks and exchanges. As trading volumes continue to rise, and trading data volumes soar even higher in the wake of Regulation NMS in the US and the Markets in Financial Instruments Directive (MiFID) in Europe, firms will need better infrastructure and storage capabilities to ensure that they have the ability to keep and maintain all the data they need, especially as clients and regulators often require that historical data be quickly and easily accessible.

“The biggest culprit of datacenter sprawl now is not compute power; it’s storage,” says former Wachovia technologist Tony Bishop. “There are so many new types of content, such as video and rich graphics, that eat up storage, but there is no cohesive strategy on how to manage it.” Rich data naturally takes up more storage space, but market data alone is presenting a big enough challenge, aptly illustrated by NYSE’s trade and quote (TAQ) data, which is produced on a daily basis and consumed heavily by its customers.

In 1998, NYSE produced 2 million to 5 million trade messages per day and now, 10 years later, the exchange produces more than 650 million trade messages on its busiest days. But in that same time period, the available speed of central processing units (CPUs) has gone up by only a few factors, pointing to the need for additional capacity to manage the data. The NYSE and other exchanges are likely on course to produce 1 billion trade messages per day in the near future and as that unfolds, the banks that receive that data need urgently to get their houses in order and build better storage capabilities.

DATABASE ISSUES
Part of the solution to the storage problem is to install a suitable database—a fact well-known to time-series database vendor Kx Systems, which has been deploying its column storage database to some of the largest sell-side banks over the past 15 years. JPMorgan, Goldman Sachs, Merrill Lynch, Morgan Stanley, Deutsche Bank and NYSE Euronext itself all use the Kdb+ database for its ability to store and quickly access huge amounts of real-time and historical data. Kdb+ uses the K programming language, created in 1993 by former UBS and Morgan Stanley programmer Arthur Whitney, now founder and president of Kx Systems.

“In a traditional database, the data is spread all over the hard disk and if you want to get six months of data back on a single stock, it could take several minutes—too long to do events processing or pre-trade risk,” says Simon Garland, chief strategist at Kx Systems. Kdb+ provides one single application for streaming and historical data. The database stores data in column format and combines that with the Q programming language, which allows very high query performance with short processing times, especially on the latest multicore hardware.

“The main thing that has drawn the banks to Kdb+ is the speed of data capture and speed of access,” says Charles Skelton, an independent consultant who has been involved in Kdb+ implementations at three large banks. “It often starts with one desk and then grows organically within an organization—I’ve seen people falling off their chairs at how quickly the data can be processed.”

TECH SOLUTIONS
Aside from databases, other new technology vendors are busy preparing the ground as banks seek to solve their storage problems. Paul Borrill, founder and president of Palo Alto, Calif.-based Replicus, has been involved in the data storage industry with Sun Microsystems and Veritas for many years and says the financial services industry is gradually waking up to the challenge. He has spent the last three years developing simple-to-use storage technology for Replicus. He hopes to roll it out to clients later this year.

Borrill’s aim is to create scalable storage systems that manage themselves and avoid data duplication, without the need for human intervention. “CIOs today spend 20 percent of their budget on new hardware and software each year, with the remaining 80 percent spent on administrative costs and human resources,” he says. “That ratio continues to get worse with no end in sight and I believe that represents a true crisis in the storage industry.” It is a crisis that the financial services industry cannot afford to ignore.