Recent crises have drummed home that banks need to calculate risk exposures in as close to real time as possible. To do that, risk managers need to process huge amounts of data, but current systems often lack the capability. How can banks address this deficiency? By Clive Davidson

Traders were stunned on May 6 when the Dow Jones Industrial Average fell by nearly 1,000 points before rebounding – its biggest intra-day loss since 1987. This price action was accompanied by a surge in the number of trades and quotes on the New York Stock Exchange (NYSE) – close to 1.2 billion were recorded, compared with a daily average of between 400 million and 600 million.

Regulators and academics are now trying to work out exactly why the Dow moved so far so fast. Blame was initially put on a so-called fat-finger trader, but much of the attention has since focused on the role of algorithmic trading strategies. For some, though, the events of May 6 have highlighted a seemingly more mundane issue – the capacity to monitor and record data.

Sudden spikes in trade data are becoming increasingly common: in October 2008, the NYSE experienced a similar jump to 1.1 billion trades and quotes. Crucially, these events pose enormous problems for key systems within banks and, in particular, those on which traders and risk managers rely. These systems were typically designed for more moderate and orderly increases in trading activity, but are now being outpaced by the markets.

Coping with a sudden surge of data inputs – critical for intra-day risk management – requires a radical overhaul of risk system design and a new generation of technology. Above all, it requires a complete rethink of how to handle data and a reappraisal of the conventional notion of what a database is.

“NYSE trades and quotes are not even one of the biggest data sources banks must deal with. However, it does illustrate a trend in the markets, which is that data levels might be relatively consistent for a number of years and then jump right up. This means application designers can’t just build an extra 10% capacity into their systems and hope for the best, because there might be a doubling of data volumes from one day to the next,” says Simon Garland, chief strategist at California-based high-performance database supplier Kx Systems.

Getting the design of data management capabilities right is therefore critical if risk management systems are to keep up with financial markets. Without the capability to handle a massive increase in data, risk systems may struggle to function. “Recently, expectations of risk management systems have escalated to the point where many of the traditional technology offerings are at best straining or have buckled and are incapable of delivering on requirements,” says Guy Pitman, a consultant working for capital markets technology services company First Derivatives in Newry, Northern Ireland.

The current generation of derivatives trading and risk management systems are mostly built on a classic three-tier business application model, with a separation between the user interface (screens for traders or risk managers), business processing (pricing or value-at-risk calculations) and data (transaction or market information). For a system of any size, the tiers will sit on physically separate hardware. Databases are usually relational and row-based in architecture (all the information relating to a record – for instance, a trade – is held in a row) and reside on disks.

“Too often, banks pay insufficient attention to the data and choices of technology when they are considering risk projects”

Andrew Bond, Oracle
There are many reasons why this model struggles with today’s markets and is not viable for managing risk intra-day. First, it presumes a dedicated database for each application. If a bank opts for a unique best-of-breed system for each asset class or risk factor, it will acquire separate databases for every application — all likely to handle data in a slightly different way. This makes it difficult to integrate data for aggregation and obtain an enterprise view of risk.

"One of the problems with this multiple best-of-breed system approach is that a lot of data gets duplicated in these silo databases," says Stuart Grant, London-based business development manager for financial services for Europe, the Middle East and Africa at Sybase, a California-headquartered database supplier. "The data starts off the same, but then starts to get out of sync as each business application cleans or enriches the data locally. This can lead to big cost or time overheads because results have to be reconciled."

The second problem with the traditional approach is that data must be transported from the hardware on which the database sits to the processing engine and back again. This takes time, especially as data volumes balloon in size. In fact, these data spikes are difficult to cope with unless the bank has a very large database server — which would be running most of the time at a fraction of its capacity. "Banks don’t want to buy an expensive server to handle a huge spike one Thursday and then have it sit idle for the rest of the year," says Garland at Kx Systems.

There is also the issue of the design of the database itself. Traditional row-based relational databases have proved very effective for storing large volumes of records in an organised way and making them available for querying, but this is not optimal for managing risk intra-day, say developers.

"Row-based relational database technologies are good for accessing small amounts of data fast enough for human interaction. But in the risk arena, Monte Carlo simulation performance requires a much faster data access paradigm," says Pitman of First Derivatives.

So how should banks tackle these issues? There is a growing consensus that intra-day risk management requires a whole new architecture and the use of the very latest data management and processing technologies.

"Too often, banks pay insufficient attention to the data and choices of technology when they are considering risk projects," says Andrew Bond, core technology director, performance and availability, for Europe, the Middle East and Africa for California-based technology supplier Oracle. Many banks have little idea of the volumes of data they need to deal with currently, let alone in the future. There is also a tendency to use the systems at hand, rather than look for newer technology that may be more appropriate, adds Bond: "Banks say ‘we have a hammer so we are going to use it’. However, for part of the solution, a database might not be the right choice. Banks need to look at the problem as a whole, break it down into its constituent parts and design a solution that is a combination of a functional architecture and a technical architecture."

One of the first challenges is to create a single database for risk. This means collecting, cleaning and standardising risk information from the various databases — a task many banks are working on. The next challenge is to reduce the latency in accessing the data. This can be achieved by removing the data from relatively slow conventional disk drives, where information has to be fetched and carried to the processing algorithms, and placing it in computer memory where it is instantly accessible. The steady development of silicon chips means it is now feasible to hold an entire database in memory — a machine can now easily hold a day’s worth of NYSE trades and quotes in memory, with top-end computers offering a terabyte of memory.

Another way to reduce the time lag is to take the analytics out of its separate processing engine and move it into the database. This is an inversion of the conventional systems architecture.

"Traditionally, a database was a bucket in which you kept data, and analysis was done in the application. Therefore, you had to move the data to the analytics engine. Now you can embed the analytics engine in the database and dramatically reduce the analysis time, as well as increase the volume of data you can put through the analytics," says Grant.

Some argue it is essential to take this approach as data volumes increase. "If the analytics are sitting there with the data, you can jump immediately to whatever data you need for your calculation, whereas if you have to ship 100 million records from a database located somewhere else to your analytics, nothing is going to get done instantaneously," says Garland at Kx Systems.

The type of database matters too. First Derivatives’ Pitman favours one that stores data in columns rather than rows, because it tends to be faster at handling the relatively homogeneous and repetitive information of trading and risk.

"Research has shown the most efficient model for a risk or analytical database is an in-memory column store model," he says. Column databases lend themselves to parallel processing approaches and associated programming languages that can show ‘orders of magnitude perform-
advance improvements over traditional databases,” Pitman claims.

Moving the analytics alongside the data and using column databases should speed up the analytical processing, but these solutions assume the data is fixed (even though the database may be constantly updated). Where markets are moving very rapidly, the aim might not be to capture and hold the data for processing, but to monitor it for certain events or conditions. In this case, a different approach can be more efficient: complex event processing (CEP) technology.

CEP tools assume a moving stream of data and use analytics to monitor the flow for events, whether they be price movements, shifts in correlations or market anomalies. Users can specify what they want the analytics to monitor, and what action should be taken when it identifies an event – for example, executing a trade when a price falls below a certain level.

CEP technology is commonly used in algorithmic trading but is now becoming part of the toolbox for intra-day risk systems, say vendors. “Real-time intra-day risk requires delivery of real-time rates and trades that can best be done using complex event processing algorithms linked to real-time databases,” says Pitman.

“Until now, when a bank reached the point where a task had become too big to run on one machine, it would have to turn to a grid with a number of servers all communicating with one another and sharing work, which required a team of experts to keep running. The new technology enables you to put the whole task on one machine with many CPUs and vast amounts of memory. So despite the increase in data volumes, there is a way to keep it simple. That it doesn’t have to get terrifyingly complex is unexpected good news,” he says.

“You can’t just rip out the entire risk management infrastructure in one go and replace it because it will compromise large, complex and critical systems’
Stuart Grant, Sybase

These steps are critical to achieving the kind of performance necessary for intra-day risk analysis, say vendors. But there is a further challenge – how to create a system that will meet market and business demands, including exponential growth of a business line or the overnight doubling of a data source. “Grid is the only answer to the challenge of scalability,” says Oracle’s Bond. In computing terms, a grid is a network of components across which processing or data can be distributed, and which can be scaled up by adding further components. The components can be dedicated central processing units (CPUs) or memory chips, servers in a data centre, or idle office computers.

London-based Hyper Rig has developed an intra-day risk architecture that uses what it calls an intelligent grid for distributing risk data and analytics. The grid is large enough to meet anticipated requirements, with an area configured for the normal processing load and the rest available for the bank’s other IT tasks. As demand increases, intelligent control software automatically scales back the non-priority IT tasks and commanders the free grid nodes for risk processing. This enables the grid to cope with sudden spikes in data or market crises that demand extra analysis, claims Hyper Rig founder Michael Coleman.

Meanwhile, a new generation of servers is emerging with large numbers of CPUs and high-capacity memory built in, and these single machines are easier to manage than computer grids linking multiple machines, says Garland of Kx Systems. “Until now, when a bank reached the point where a task had become too big to run on one machine, it would have to turn to a grid with a number of servers all communicating with one another and sharing work, which required a team of experts to keep running. The new technology enables you to put the whole task on one machine with many CPUs and vast amounts of memory. So despite the increase in data volumes, there is a way to keep it simple. That it doesn’t have to get terrifyingly complex is unexpected good news,” he says.

“You can’t just rip out the entire risk management infrastructure in one go and replace it because it will compromise large, complex and critical systems’
Stuart Grant, Sybase

These scalability and performance advances mean banks can begin to move from overnight batch processing of data for enterprise risk analysis to something approaching real-time analysis. It may not yet be possible to achieve the Holy Grail of continuous, instantaneous calculation of exposures, but developers believe that by using the new hardware and software technologies, it is possible to obtain risk updates at quite short intervals.

“If you are doing intra-day risk analysis, the data has to be complete – you can’t have anything missing otherwise the results will be misleading,” says Hyper Rig’s Coleman. “But it is possible to manage the data down to fairly fine-grained chunks of five to 10 minutes. The chunk may be five minutes behind the markets, but it will be consistent and complete. Knowing your exposure as of five minutes ago is still a lot better than knowing what it was yesterday.”

The other function of databases is to act as long-term stores of information. Regulations such as Basel II and the European Union’s Markets in Financial Instruments Directive require banks to keep information for long periods (up to five years), so firms have no option but to build large archives. The new architectures for intra-day risk allow data to be copied or moved from memory to data management environments that can use specialised hardware and techniques, such as compression and zipping, to reduce both the storage media requirements of archiving and the data transfer rates when there are queries on the historical data.

Although these new high-performance data and processing architectures offer the promise of near real-time risk analysis, few banks have greenfield sites in terms of technology. Most have already invested heavily in trading and risk systems they have fine-tuned to their operations. “You can’t just rip out the entire risk management infrastructure in one go and replace it because it will comprise large, complex and critical systems,” says Grant of Sybase. He advocates a component approach, whereby the bank migrates one section of its risk management at a time. “This offers a safe route that will also show some benefit to the business in the short term, which is important today,” he adds.

Meanwhile, Pitman points out any new intra-day risk architecture would not necessarily replace all the functions of existing databases and systems. “A relatively small amount of the information held in existing treasury systems is relevant for risk management,” he says. The rest, covering aspects such as settlement and accounting functions, is adequately handled by current technology. With the proviso that a bank’s existing systems are reasonably up to date and have the interfaces that allow for integration and data access, then “it is possible to slide a new system into an existing technology infrastructure without getting rid of old systems”, agrees Coleman.

Events such as the collapse of Lehman Brothers and the Greek debt crisis have driven home to banks the need to be able to calculate enterprise-wide exposures at short notice. Given the speed and data volumes of today’s markets, it is a big task. However, the new IT architectures may begin to help.