

A typical investment bank has a plethora of data – both structured and unstructured – scattered across the enterprise. There are data feeds from exchanges, third party vendors such as Reuters, Bloomberg and Thomson and internal data feeds from historical databases containing reference or transaction data. “Trading increasingly requires more and more data, yet many institutions do not have the necessary infrastructure in place to manage it successfully,” observes Brian Sentance, commercial director, Xenomorph, a provider of enterprise-wide data management and analysis solutions. Or as one vendor puts it, banks are literally ‘choking’ on data, and as more trades are conducted electronically the amount of data banks have to contend with is showing no signs of abating. Added to that is the increasing regulatory imperative to not only store historical data but to be able to detect fraud or transactions that exceed acceptable risk limits before they turn into catastrophic events.

“Part of the problem is being able to turn data into data bytes of information that are useful for the eyeball and which enable you to make intelligent decisions,” says Andrew Miller, managing director, Arcontech. Sounds simple. After all isn’t data information? Well not if it is coming at you at unrivalled speeds (real-time tick data) or if it is stored away in historical databases scattered across the enterprise. There is also the question of how reliable that data is. Can banks ensure the data they are using, which forms the basis for trading strategies



# The data conundrum

As more trades are executed electronically, regulators demand higher levels of data capture and traders seek competitive advantage, data-base technologies are trying to keep up with the need for speed.

and risk analysis, is quality data? “As trades are done electronically and as people expand their trading algorithms in terms of the number of different trading algorithms they are using, they have to get the right data,” says Randy Grossman, a senior analyst with Financial Insights. “If reference data is not correct then algorithms will not run properly.”

Financial services firms are under pressure from competitors, regulators and shareholders to derive business

value from this data, to analyse and interpret it in such a way that it will give them a competitive edge. And sometimes this needs to be done in a matter of minutes, rather than next day or a month later. “Implementation of Volume Weighted Average Price or index arbitrage trading strategies need to operate in real time,” says Dr John Bates, CTO of Apama, a provider of trading strategy management technologies. California-based ObjectStore, a provider of real-time data management products, cites the

example of an investment bank that wants to execute mathematical trading strategies on every stock tick in real-time in an effort to identify trading opportunities potentially worth millions of dollars.

“The time of the inactive asset manager is gone,” says Willem De Geer, CEO, Panopticon, a Swedish-based company, which develops heat-mapping solutions for rapid analysis and visualisation of financial and market data. “If you manage a portfolio of stocks it is about the feeling of control. It isn’t the data so much that is adding value, but how data affects other data.” In other words, the complex interactions and correlations between different data sets, patterns that may be difficult to detect using traditional data visualisation tools, which do not necessarily operate in real time.

### Technological limitations

But the increasing need for speed raises questions about the way firms have traditionally stored and retrieved data. Simon Garland, senior vice president, technology, Kx Systems, which provides solutions for capturing and analysing data in real time, says the speed with which data needs to be handled means a decision needs to be made about what data to keep. “A lot of people still have the view that they cannot store everything and they should just keep consolidated versions. If you are storing the average price per trade for lots of stocks you don’t need any time series manipulation. You

could put it in an Excel spreadsheet, but as soon as someone wanted to look at that data intraday, you would need to launch time series, which you couldn’t do with traditional databases.”

In a White Paper on the subject, Kx says not only is it difficult to manage the enormous amounts of data flooding the enterprise, “it is nearly impossible to perform instantaneous comparisons of streaming data to historical patterns – a requirement for making real-time business decisions.” It cites the example of a real-time event such as credit card fraud, which needs to be compared with historical data to detect patterns.

Barry Raskin, president, Telekurs Financial, a provider of real-time securities data from more than 200 exchanges, says the sheer volume of data required today is throwing up processing issues. “How much of the real-time data do you need to provide? Do you need to provide every single last tick, every piece of data that comes

from an exchange?” Raskin says the bandwidth required to process the increasing volumes of data is challenging. “We are working directly with a lot of exchanges to figure out whether they are providing clients with the information they really need.” Storage capacity is not an inexpensive commodity, and with trade and transaction data multiplying at a phenomenal rate, the question is more what to store and how to store it for quick and easy retrieval.

Yet, in an increasingly competitive data driven world, banks and vendors are being asked to stretch the capacity of databases to their uppermost limits. “The data infrastructure in general in capital markets is sub-optimal,” says Gene Kim, senior analyst, capital markets, Financial Insights. There are many different kinds of database – relational, time series, object-oriented. In this competitive climate, vendors tend to play up the competitive advantages of one approach over

### Dr John Bates, Apama

Implementation of Volume Weighted Average Price (VWAP) or index arbitrage trading strategies need to operate in real time.



another. An example of this is relational databases, which were invented in the 1970s versus object-oriented databases. Relational databases represent information similar to values in a table. A Structured Query Language (SQL) is used to define, manipulate and query data in a relational database. An object-oriented database makes database objects appear as programming languages (typically Java or C++), and is more suited to projects where there is a business need for high performance on complex data.

## A new 'paradigm'

Kx Systems' technology is based on a 'high performance' relational database. Xenomorph, with its background in time series analysis, does not deny that relational database technology is proven and successful. However, Sentance says the relational model does have limitations. "The relational data model is designed as a generic solution for the storage of generic data. Given that it is so general, SQL access cannot compete with database technology built specifically for specialist, more complex types of data, such as time series, curves and matrices," he says.

The London-based data vendor claims its technology can deliver retrieval speeds over 100 times faster than relational databases for time-related querying. Xenomorph has recently upgraded its database engine to cope with high volumes of time series and tick data. "Storing the data is also only

a small part of the problem," Sentance explains, adding that in financial markets firms required access to data frequency conversion, data source preferences, missing data rules and interpolation methodologies, which he says cannot be expressed via standard SQL statements.

In its White Paper, Kx says SQL is slow when it comes to analysing 'time-ordered' data, but Garland believes that relational databases can be made to perform for today's real-time demands. "Some people think the only way forward is to have a totally new paradigm, but if you look at a relational database, there is more under the covers." Kx's relational database uses a sophisticated programming language to speed up the management and analysis of both real time and historical data.

"Companies want to be able to access historical data in exactly the same way as real-time feeds," says Garland. "The shift between my real time and historical feeds should be transparent."

ObjectStore claims to process 'spiralling' volumes of data 100 to 1,000 times faster than traditional relational databases. It says its object-relational data-caching technology, enhanced by its recent acquisition of Persistence Software, reduces the latency times of banks' back-end data systems. It has also released Trading Accelerator, which allows traders to perform 'what-if' queries on incoming data or real-time trading strategies such as VWAP analysis. Accelerator includes a 'tick' database, which captures market data

feeds at high speeds then compares it with historical and reference data values.

Whilst a robust data infrastructure is needed to capture large volumes of tick data for algorithmic trading, Financial Insights' Grossman contests whether some of these trading strategies need to operate in a real-time environment. "People who have written these algorithms say you are not going to see that much intraday or real-time processing as the relationships between the various algorithms don't change that much over the course of the day." Perhaps it is more a perception of competitive advantage. If firms can detect trading patterns quickly, as well as being able to change strategies 'on-the-fly', then the belief is that one institution may be able to gain a competitive advantage over another.

Yet, Kim of Financial Insights believes vendors are doing themselves a 'disservice' by making grandiose promises about speed. "I don't think you can say this [relational databases] is better than the other [object-oriented]," he says. "It depends what you're trying to achieve." Asked to differentiate amongst the various database vendors, Kim says the only markedly different approach he has come across is that of start-up Avaki, which uses grid technology to access structured and unstructured data across an organisation in real time. "The Holy Grail of course is that the enterprise can run on a single database," he says, but short of asking the major



## Willem De Geer, Panopticon

The time of the inactive asset manager is gone. If you manage a portfolio of stocks it is about the feeling of control.

investment banks to rip out their existing data infrastructure, the next best thing, he says may be data federation – the ability to access multiple data sources within a single query. “I don’t think the industry can continue to do what it is doing now, historically having lots of different databases and data in different formats,” Kim continues.

### Analysis on-the-fly

Whilst databases themselves are being upgraded and modified to perform in today’s real-time environment, tools for analysing data are also undergoing dramatic change. Apama’s Event Manager is not a database, but a ‘funnel’ inserted in the data flow. “It just treats historical databases as another data feed,” explains Peter Beard, Apama’s CEO.

Apama does not replace database technology, but tries to overcome the limitations of traditional database management systems and rules engines by allowing traders to detect and respond to complex data patterns more quickly. “Other technology vendors have tried to address this problem by storing and

indexing data,” Dr Bates of Apama explains, but he says this approach is constrained by the fact that every time a trader wants to act on a new piece of information, they have to re-write the index, which makes it difficult to keep up with ever-changing data.

Apama’s Event Manager sits on top of the data flow and works on the basis of user-defined trading strategies. “The data is then flowed across the strategy,” Dr Bates explains. “Data flows in and sticks to the strategy and takes the necessary action.” The Event Manager generates an alert, which indicates an action has been taken (a trade executed) or an analytic computed. “Traders can identify threats or trading opportunities in sub-milliseconds,” says Dr Bates, “and create new strategies on-the-fly in minutes as opposed to several weeks.”

### Colour your world

With the emphasis on timely analysis and securing competitive advantage, Swedish-based Panopticon leverages heat-mapping technology, which uses different colours to present market information in a more easily

digestible and contextual format. So for example, in the case of stock prices, blue could mean the price of a stock has gone up, red meaning the price has gone down. It also allows for quantitative comparisons in terms of broker’s buy and sell recommendations (blue for buy, red for sell and white for hold). Market or financial data can be broken down into different data sets and users can view the data over different time periods (a week, a month or a year). Explaining the benefits of the technology, Panopticon’s De Geer says, “If I looked at a portfolio in numerical format I wouldn’t see the context. But this puts data quickly into context. With real-time heat mapping, as soon as the tick has happened you can spot it.”

Telekurs’ Raskin believes that whilst databases will continue to grow to handle increasing volumes of data, the emphasis will be on data management tools that help sort through this information in a timely fashion. “There’s a need for some processing somewhere. Everyone is looking for a competitive advantage, so I think it will move more towards the business side and away from the technology side,” he says. //