Spreading wings

Data management vendor Xenomorph is stepping up its global presence. We speak to **Brian Sentance**, CEO of Xenomorph about the company's latest developments and recent collaborations

Could you tell us what the company has been busy with in 2007?
One of the biggest areas for us has been data validation and data cleansing. This has been driven primarily by the regulators, who have been saying that the data going into the risk management models are of poor quality. Many of the risk management vendors assume the world of data is perfect and complete, which is a very big source of failure within some of the risk management projects.

Feedback from our clients and prospects indicates that risk managers (and traders) can spend 30 to 50 percent of their time cleaning up and validating data, so that's been a strong focus for us with regard to the number of clients we have in data management for risk. In other words, we want risk managers to spend more time managing risk and less time on managing data.

Given this background, the next release of our data management solution, TimeScape version 3.2, has a number of enhancements related to this focus on data quality.

One particular feature is having more than one person involved in the validation process. Basically, one person makes the change, while the system forces another person to approve the change. The market refers to this as "Four Eyes" validation, which we have generalized to "*N* Eyes." Well, why not? Someone will ask for it. On a similar note, as several data vendors tighten up the area of data access, we have also introduced a very fine granular permissioning within our systems. For example, a particular group of risk managers or quants can access Bloomberg's data but not data from Reuters, or could look at data for one asset class but not for another.

Another functionality is the ability to capture real-time data and snap it on a global basis, in order to ensure the consistency of data across the markets. Assuming that you use end-of-day prices across the globe, if you look at the statistical behavior of that data, you are out of step all the time. Hence, the actual risk calculations can be very different from if you just look at your exposure simultaneously, for example, at 4:30 pm Hong Kong time. So it is important to be able to snap real-time data consistently across a large number of instruments due to the statistical effects it has on data and the calculations.

We have also done some performance enhancements on tick and time series data. In particular, we have been collaborating with Microsoft and are currently benchmarking some technology we have developed with them. Basically, we are moving towards implementing our proprietary database within Microsoft's SQL server, instead of having our database reside on disk. The motivation behind undertaking this work has been to deliver what traders and risk managers want, which are proprietary performance levels, while at the same time delivering the kind of database standards-based approach that technologists would prefer. So far, we are benchmarking this at around one terabyte of market data, which is about 36 billion tick prices! Also, we're working with Microsoft to address larger benchmarks, into the tens of terabytes. So far the results are very positive.

What has proved particularly challenging for the company in 2007?

Data volumes with the tick data is something that is constantly challenging. So enhancing our systems to support greater volume and data storage, and being able to analyze it in real time, has been quite an important part of what we've been doing. We realized that some of our competitors have a technical architecture where data does not get written to disk until overnight on a batch. Instead, it is written to another database because there is just so much tick data. So they have an in-memory database and a hard disk persistent database. It seems illogical to us because it makes it harder to combine and analyze today's data with yesterday or last year's data. Therefore, one of the technical challenges we have been working on is making sure that what we've done in real time and tick data is consistent and easy to use, whether the data has just come in (real time), is today's data, or is tick data from several days or years ago. Why should users have to care about how data is stored? All they want is the data they need as quickly as possible!

Could you elaborate a bit on collaborations with Microsoft?

There are three areas of collaboration.

We designed our database differently from the relational database to allow very fast access to historical and time series data. Quite simply, the data is organized in a way that allows you to jump to a particular instrument very quickly and from there, you are able to jump to a particular day, year or time very quickly.

A few years ago, a client and partner suggested that instead of hosting that binary data on disk, we should consider if we could host it within the SQL server. Therefore, at the end of 2006, we worked with Microsoft to try different ways to achieve that, combining our XDB technology inside the SQL server. Following the success of that, we spent another month implementing a particular way of doing it with the Microsoft UK DPE team (DPE stands for Developer Platform Evangelist; Microsoft loves acronyms far too much in my opinion!). All the coding has been completed now, and we are at the testing and benchmarking phase.

The results we find using the SQL host have been very encouraging, even for very large databases. We find access time to be very comparable to what we already achieved with XDB and unlike a normal relational implementation, the performance does not degrade as database size increases. Although we are getting the results that we want, as mentioned earlier we want to push the limits by further testing it to the full with very large databases. We will be launching this with Microsoft in the new year and there will be events both in the UK and across Europe.

The other area we are working on is Excel Services. We are looking at taking spreadsheet operations that people already have to see how best we can manage that and add value to it by combining that with our database technology. I think the Excel Services is already a very useful tool for the quant community, but this is more or less the first version of Microsoft's attempt to turn Excel into a server calculation engine. Initial feedback was that you still have to do a lot of programming to use it. I am sure that this is one of the key areas Microsoft is looking at. Another area that may take some time to sort out is the fact that you cannot use real-time data on Excel Services yet, which obviously limits how easily desktop spreadsheets can be moved server-side and centralized. Many interesting features will be added in the coming versions, which will make Excel Services more useful for the quant community, and we are working hard to align Excel Services with our own "spreadsheet in a database" technology, TimeScape SpreadSheet Inside.

A closely related development is what Microsoft is doing with its entry into the calculation clustering market with its HPC clustering technology. We are looking to use that technology to address two areas.

The first is to give greater scalability and reliability to instrument pricing for very large portfolios of derivatives. This application of HPC is aligned with our thoughts that there is still a lot to do in easily combining data with derivatives analytics, thus extending data management into the new field of analytics management.

Moving away from the more complex products for a moment, the second area concerns real-time tick data management where there are other challenges, notably being able to handle large volumes of real-time data in-memory. In this problem space, clustering and data grid technology offers the potential to distribute database storage and cacheing across a cluster, which will become increasingly important as the data volumes involved are enormous and ever increasing.

That said, I guess as soon as you raise the bar and provide the capacity to deal with the problem, the bucket gets filled up very quickly. So because you can do it, people want to do it; therefore, they get a new limit, which may be extraordinarily large or challenging, relative to what they have done in the past.

A quick example of this comes from a friend who works on network infrastructure at one of the UK investment bank. One of the traders wanted to send a message to Tokyo and receive acknowledgment that the message was received. So he gave him the timing of the round trip to Tokyo from London and back, which was a few milliseconds. The trader said that this wasn't fast enough. He explained to the trader that if you take the distance to Tokyo, multiply that by two, take the speed of light and do the math, you actually arrive at a time very close to the particular time this will take to happen. Yet the trader wanted to break the laws of physics to achieve what he wanted from a business perspective!

Yet the trader wanted to break the laws of physics to achieve what he wanted from a business perspective!

So, you can end up with technology colliding with business requirements. Looking at tick data, with regard to algorithmic trading and the introduction of Mifid, with the increase in trading locations, the fragmentation of liquidity in Europe, and following what is happening in the US, trades' transaction sizes will go

It is a "we've got the data, now what?" kind of problem. Hopefully, the clustering approach will help in this respect

down. However, there are many more transactions across more locations as people try to hide what they are doing, so the actual data volume will go nuts, partly because of regulations and partly because it is possible and traders want to do it.

With regard to the subprime crisis and the credit crunch, how have issues such as these affected the company?
On a negative basis, all the vendors in the market face the possibility of having projects delayed. However, I also think it's a real opportunity for vendors involved in risk management and data management to actually put forward the benefit of having greater transparency around accessing data, such as credit derivatives data for all the different players - the quant, the risk managers, trader, and everyone involved. We will be able to market the need for our product.

I find it fascinating how things have become more connected and not necessarily in straightforward ways, such as what's happened to Northern rock, who weren't necessarily involved in bad lending. Another aspect for quants is having a theoretical price for an instrument (using the most advanced quant techniques in the world), which counts for absolutely zip in the market if no one wants to buy the security.

I recently heard that E*Trade's share prices have fallen badly because they have been lending into subprime. I thought it was an execution company? The story is that they opened up bank accounts to support execution and ended up offering mortgage products. So who's next? Obviously, some of the things concerning ABN Amro, RBS, and all the people involved in that deal, the subprime crisis obviously made the cost of financings a lot higher and, as a result ,introduced greater delays into many mergers and acquisitions, which is another way projects can be put on hold. There are both positive and negative aspects to the crisis in summary, and delays in some projects, but they also represent some new opportunities and a real chance to put forward what we do with a real and topical business context.

What can we expect from Xenomorph in the new year?
We will be moving a lot of our technology to 64-bit architecture, partly driven by tick data, but also the whole industry's move to bigger chips and larger memory. We should be seeing this in the first half of the year. We are also working with the Microsoft clustering team, with their HPC Clustering technology, to both scale out the calculation loads that we have across multiple servers and to improve the level of reliability and fault tolerance in using the grid or clustering approach for data analysis.

One major problem in the industry is the massive amount of data. You can handle statistical analysis on end-of-day prices inside a spreadsheet, but if you are looking at the FX rates or a very liquid stock - for example, Japanese yen to the dollar - you might have as many as 300,000 price events per day. If you are looking at a year's worth of that data, that is a very big problem to handle. Even in the new versions of Excel, which contains one million rows, that is three days' worth of data, so that is not very much. It is a "we've got the data, now what?" kind of problem. Hopefully, the clustering approach will help in this respect.

We are also seeing a lot of interest from both the derivatives and risk management areas for the management of unstructured data. Therefore, we are looking at not only storing market data, instruments, volatility services, or pricing models, but also storing work documents that contain all the terms and conditions sheets and integrating that with their spreadsheet environment. We believe that business users don't really care about the technology; they just need to see all the data they need together and easily accessible. If you can do that, you can be more productive.

All the technology that we are developing supports the data management for risk and derivatives, and also is able to deal with new products and help our clients bring their ideas to market a bit more quickly.

What would be your hopes for the New Year?
My belief is that the banks are going to ride the credit crisis, because I believe that there is still good money to be made in other parts of the business, for example, in equity. I hope the banks will get the bad news out of the way quickly, since both markets and the economy do not react well to bad-news announcements that get worse as each quarter goes by.

From a business perspective, we are getting a lot of sales leads on a global basis and not just near our European and North American sales offices, this is actually helped by *Wilmott*. One new client is based in South America, a new region for us but with new data and financial challenges to meet.

And we're also looking to expand Asian operations within the next 12 months and we welcome any new applicants to staff this office to get in touch!

We'll also be moving to 24-hour support models to support some of the global roll-outs we have recently implemented. In summary, we're looking to complete a global presence for the company, with Asia being the final piece of the puzzle to be sorted out. Asia is again, another region presenting its own particular data management and analysis problem. We are getting a significant number of interesting projects and sales lead in this region, without actually doing any direct marketing, and this is happening similarly in South America.

I guess the world is getting smaller and smaller, which I still find surprising and very enjoyable. For instance, when I was in New York at a Xenomorph event a few weeks back, I sat down at an airport lounge for a bite to eat, and who sits down at the next table but Paul Wilmott! It is utterly bizarre that we struggle to arrange a meeting to catch up in London, but we both just happen to sit down at the same second, at the same restaurant, at an airport a long way from home. I still think the market for "micro-derivatives" is still wide open, and in this regard maybe Paul needs to price (and buy!) a "let's eat my dinner alone without interruption" option...