

## Trend Spotters

### **Finding needles in haystacks is getting a little easier, as data-visualization software for businesses matures.**

By Rick Whiting, Sept. 5, 2005

The Department of Homeland Security has a keen interest in the development of data-visualization tools that make it easier for intelligence analysts to extract key pieces of information from large volumes of structured and unstructured data. With terrorists taking advantage of the wide-open borders of the Internet and E-mail messaging groups to find recruits, raise money, and encourage jihad against the United States, the department needs a way to ferret out immediate threats as well as uncover sleeper cells and other dangers. So last year, it created the National Visualization and Analytics Center at the Pacific Northwest National Laboratory in Richland, Wash., with \$2.5 million in funding to develop technology that combines data mining and analysis with visual-presentation capabilities to detect, assess, and reduce the threat of attacks, as well as respond to them.

The road to graphically rich, texturally detailed graphics that can help the center spot threatening trends--and might someday help businesses identify undervalued stocks or locate supply-chain bottlenecks--is being charted only now. Scientists and engineers have used sophisticated data-visualization tools on workstations for decades to design cars, model molecular compounds, and simulate the effects of tornados and other weather phenomena. But using computer graphics to model something tangible like airflow over an airplane wing is one thing. Adapting that technology to create interactive visual representations of less tangible data, such as terrorist activity or stock-trade patterns, has taken longer.

That's left much of the business world in the dust, even as data from Web sites, point-of-sale systems, and IT security tools rapidly accumulates in companies' data warehouses and the speed of business ramps up. People are left to glean trends and spot patterns from eye-straining text reports, floor-to-ceiling spreadsheets, and Excel's basic 2-D and 3-D bar charts, scatter plots, and histograms.

"We're inundated with information," sums up the visualization center's director, Jim Thomas. "What we're trying to do is solve the problem of working with large volumes of often complex data." The center's work, which revolves around using spatial and graphical elements to convey more abstract data, has applications for businesses, too. The goal is to provide new tools for dealing with the huge digital libraries of the future, whether to detect terrorist threats or do genomic sequencing in the development of potentially lifesaving drugs. For its first partnership, the center is working with Stanford University researchers to develop methods to analyze network traffic patterns to graphically identify compromised systems used by consumers, businesses, and the government.

Confronting data overload, a growing number of companies are tapping into a new generation of data-visualization tools for businesses that they expect will let them make discoveries--find quality-control anomalies in a production run or the most-promising results in clinical drug trials--in minutes rather than days or weeks. Think 3-D representations of a day's worth of stock trades that appear as a mountain range with multicolored peaks and valleys signifying high and low stock prices, trading volumes, and other metrics. Some of the new tools superimpose data over geographic maps--a utility company might show electricity flow or power-substation status in the area it serves--or graphically represent a computer network or supply-chain system. And a comparatively new form of graphics called "treemaps" uses squares of varying size and color to visually represent data variables (see story, "[Treemaps Rule: Even The Marines Think They're Cool](#)"). Most of the tools that create these charts let users drill down into the details behind each one.

Pharmaceutical maker AstraZeneca plc uses data-visualization tools to better manage drug-development projects. Earlier this year, some 400 managers and employees at AstraZeneca began using Advizor Solutions Inc.'s data-visualization software to track project schedules, costs, and allocated resources. Pulling together data scattered across the company in spreadsheets and databases, the software provides color-coded bar and Gantt charts that help managers get a high-level view of the product-development process that project-management apps can't provide, project information manager David Scanlon says. "You're four clicks away from a dozen graphs" that provide snapshots of project activities, schedules, devoted resources, and costs, he says of Advizor's simplicity.

Yet the state of the art among most tools for business-data visualization doesn't hugely impress Scanlon, who has a background in genetics and molecular biology. Compared with the tools used for scientific work, visual-analysis tools on the business side are "clumsy and pretty rudimentary," providing simple dashboards rather than real interactivity.



American Water uses data-visualization graphics like this to aggregate data from security systems.

Chalk up the less-evolved state of the market to a few factors, including a lack of communication between the developers of data-analysis software and data-mining software, which are tailored for sophisticated users such as statistical analysts, and the vendors of early data-visualization tools. Startups, such as FYI, OmniViz, Spotfire, and Tableau Software, are bridging that gap.

OmniViz is a spin-off from the work done at the Pacific Northwest National Lab. Tableau got its start at Stanford University's department of computer science when the company's co-founder and chief technology officer, Pat Hanrahan, a Stanford professor--and now the researcher leading the university's work with the National Visualization and Analytics Center--was approached by the Department of Defense to develop a way to help it get meaningful data from a variety of databases.

Spotfire, which received seed money from In-Q-Tel, the CIA's technology incubator, counts among customers of its DecisionSite visual-analysis tool ChevronTexaco Corp. The oil and gas company uses DecisionSite to produce charts from geological testing data, which can take up as much as 200 columns and 30,000 rows on a spreadsheet, and well-production costs that help identify sites with greater potential for petroleum reserves. Now Spotfire has developed a tool built on DecisionSite that uncovers patterns in E-mail messages and displays the results in color-coded grids. The National Security Agency is using it to help identify terrorist messages. The software also has potential text-mining applications for businesses. The situation also is improving as vendors of scientific-analysis software, such as SAS Institute Inc., and dashboard software vendors, such as Corda Technologies Inc. and Cognos Inc., turn their own expertise to data visualization. There's greater opportunity for these vendors in the business market than there was a few years ago, now that the average desktop PC has enough power to massage massive amounts of data into comprehensible graphical images. For less than \$5,000 a seat (possibly quite a bit less, for small-scale deployments), including the cost of a PC, most businesses can have access to visualization technology.

Downsizing the cost is important to budget-conscious organizations such as the Louisiana Office of Family Support, which manages the federal food-stamp program in the state. The department overlays data from Information Builders Inc.'s WebFocus business-intelligence tool onto mapping software from ESRI Inc. to help spot transactions that don't follow normal patterns and might indicate fraud, such as someone cashing food stamps at a store far from home. "You can see trends that you can't see in a report. Some of this stuff just pops out at you," says Sherwood LeMoine, a consultant in the Louisiana office.

American Water, a subsidiary of RWE AG, has found that data-visualization tools help it discover anomalies of another sort. The company, which provides water service for 18 million customers, has more than 4,000 sensors monitoring its IT network for attacks and intrusions and generating more than 100,000 warnings every day. In that environment, it's a challenge for security director Bruce Larson to sort out the real threats. "Any time you have a lot of information with the potential for a lot of false-positives and a requirement for rapid response, you need data visualization," he says. American Water aggregates data from its security systems and presents it in real time using software from Advizor to display charts, such as one that provides a representation of the source and target IP addresses of network attacks.



Data visualization helps American Water find the real threats to its systems, security director Larson says. Photo by Chris Mueller/Redux Pictures

Still, most data-visualization tools don't yet let users analyze real-time data, such as that generated by a network of radio-frequency identification sensors or data processed by a streaming data-analysis engine. One exception to this is data-visualization software from Visualize Inc., which Nasdaq uses to provide executives of its listed companies with real-time pricing and trading-volume data that can be converted into interactive colored charts. Generally, data quality and integrating data from many sources to support data-visualization applications remain knotty problems pre-empting real-time analysis.

"The shortcomings of data visualization as an interface to databases has been one of my pet peeves for many years," says Michael Stonebraker, an icon in the database world and CTO at StreamBase Systems Inc., which makes streaming-data-analysis engines. In the mid-1990s, at his object-relational database startup, Illustra Information Technologies Inc., Stonebraker developed an early graphical data-visualization user interface called Tioga that let users "fly" through and explore visual representations of huge volumes of information.

While such advanced concepts haven't caught on, for many companies the latest tools let them work more quickly than they once did. Infinity Pharmaceuticals Inc. uses Spotfire's DecisionSite tools to monitor data on drug-experiment quality control. That information--thousands of data points per experiment--used to be collected by Infinity's informatics department, which processed the data and sent it back to scientists in spreadsheets. "It would take weeks to review in Excel," Infinity analyst Jeffrey Reed says. DecisionSite displays the data using graphics that mimic the plates and their rows of wells in which chemical compounds are tested, using color-coding to tell which wells are contaminated.

Data visualization may even have legs as the next big thing in knowledge management. Clifford Chance LLP, a large international law firm, has built a knowledge-management application that provides a graphical view of, and the links among, the firm's law practices, experts in its 29 offices, and documents. A solicitor in the asset-finance practice searching for information about aircraft leasing, for example, can follow links to locate files and people with expertise in the transportation practice area. Before building the data-visualization app two years ago, using tools from Thinkmap Inc., each practice had its own database, and the firm maintained an unwieldy taxonomy in a Microsoft Word document, says CIO Paul Greenwood. "When we printed it out, it was impenetrable."

Microsoft is thinking the same way. Internally, managers have access to a system tied to Outlook that provides a 3-D graphic of how they relate to any of the company's 60,000 employees, says George Robertson, who heads up the vendor's Visualization and Interaction Research Group and serves on the National Visualization and Analytics Center's advisory board. Microsoft is building similar capabilities into a future version of its Identity Integration Service software that IT managers will use to view intersecting hierarchies of data across multiple databases. And Windows Vista will offer visualization capabilities that show how items in a list, such as the results of an Internet search, relate to one another, Robertson says.

#### **THE UPSHOT**

---

Many businesses still use spreadsheets with hundreds of columns and thousands of rows to analyze data and spot patterns. Data-visualization software can make that information more accessible.

---

It's not perfected for the business world yet. Data quality and integration issues remain knotty problems.

---

The technology has some big names interested, including Microsoft.

It's clear that the business applications for data visualization are only starting to be explored. The Louisiana Office of Family Support plans to expand the use of its data-visualization tools to such tasks as finding the best locations for the department's offices based on client traffic and analyzing spending on child-care services. "This is really a business-management system," LeMoine says. In an increasingly complex business world, that may be just what's needed.

-- with Charles Babcock