



To remain competitive, any business must understand customers. Once you know the detailed buying patterns and demographics of customers, you can launch sales and marketing efforts into narrowly defined markets. Today those markets may be so narrow that one large bank refers to them as market segments of one.

To segment customers into slices that small, companies need data—ever larger amounts of data. In fact, corporations are acquiring data so fast that one expert estimates that mission-critical data has doubled every year for the past five years. The desires to know customers intimately and to target markets accurately are good reasons why enterprise data is overrunning the corporate landscape. Additional causes may be technological, such as distributed systems and bigger applications, or social, such as industry deregulation or litigation. Whatever the reason, it seems that executives and managers can't get enough data and tools to analyze it.

There's good news, though. Storage has never been cheaper, strategies for managing data are reasonably well tested and the tools for storing, maintaining and delivering data to end users in useful ways are proliferating. Most encouraging are reports from the front. As high as some of the hurdles are, IS professionals who accept the facts of life are able to provide business colleagues with information that contributes to revenue-generating efforts.

### It's Everywhere

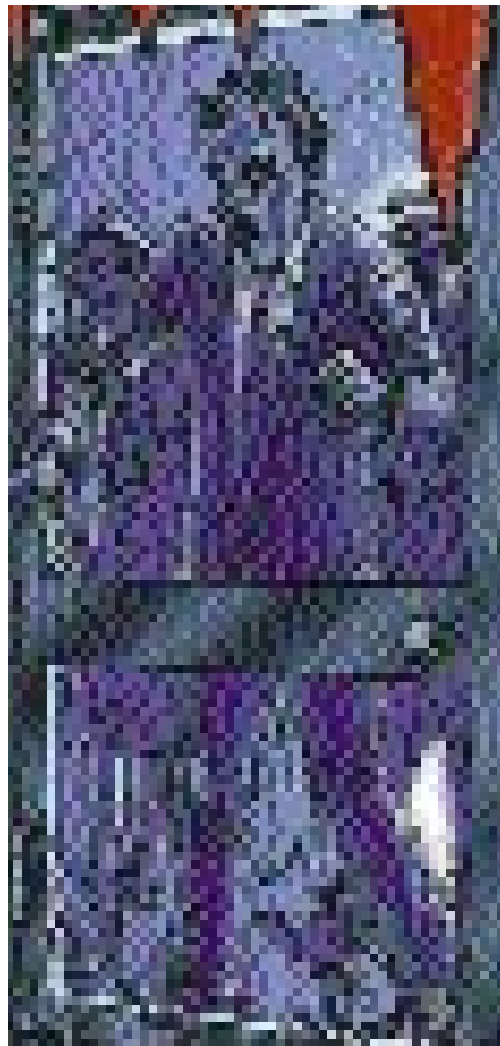
How fast is enterprise data growing? Hard figures aren't available, but anecdotal evidence is compelling. A manager of decision support systems (DSS) at a \$1 billion laboratory instruments maker says an executive information system (EIS) that he built a mere four years ago at 10MB has grown above 25GB, and it's just one of four EISs the company has.

Richard Finkelstein, president of Performance Computing, a client/server and database consulting firm in Chicago, has been building databases

# Attack of the Rampaging Data

By Bill Roberts

Enterprise MIS must handle ever-increasing amounts of data. Here are some strategies and case studies for dealing with this issue.



for 20 years. "My last big system on a mainframe 10 years ago was one gigabyte and considered huge," he says. "Now you'd say a terabyte is large. That's a thousand-fold increase."

George Ferguson, a product manager at Hewlett-Packard in Palo Alto, CA, began working in HP's data warehousing business four years ago, when most customers were asking for 10GB systems. "Today that's 100 to 150GB at least, with about 25 percent of new data warehouse business requesting more than 500GB," he says.

Not everyone requires an industrial-strength database of those dimensions. "You'll hear people talking about terabytes," says Randy Betancourt, program manager for data warehousing at SAS Institute in Cary, NC, "but, with the exception of the *Fortune* 1000, the typical size of a data warehouse is about 50GB."

Even the *Fortune* 1000 includes modest efforts. Bernard Seban, database manager for computer systems vendor Data General Corp. in Westboro, MA, has only 4GB in a data warehouse dedicated to online analytical processing (OLAP). However, this is not a comprehensive repository. "We have only four years of history in there," he says.

### Getting to Know You

As noted, the desire to know customers is what's driving data growth. Today, point-of-sale retail systems collect every piece of accessible data about a customer. Computers store it. Marketing and sales staffs demand access to it.

The goal is *micro-marketing*: getting the right promotions to the right people at the right time. "If you're spending \$50 million on a direct-mail campaign and you can identify the 10 regions that will give you the biggest payback, you can target it that way," says Michael Guidry, director of corporate applications development for Management Science Associates, a market research and applied technology firm in Pittsburgh that works with various vertical markets, including consumer packaged goods companies. In that niche, Guidry says, data is flowing into the enterprise from points of

sale; store-level marketing studies that rival the U.S. Census in detail and breadth; syndicated information; and other sources.

It has gotten to the point that, in retailing, the sale itself may no longer be the most important thing. "Sometimes the most valuable commodity that comes out of the exchange is the demographics you capture from it," says Betancourt of SAS. "There's a keen desire to capture as much information as possible about that transaction so you can go back and sell to [that customer again]."

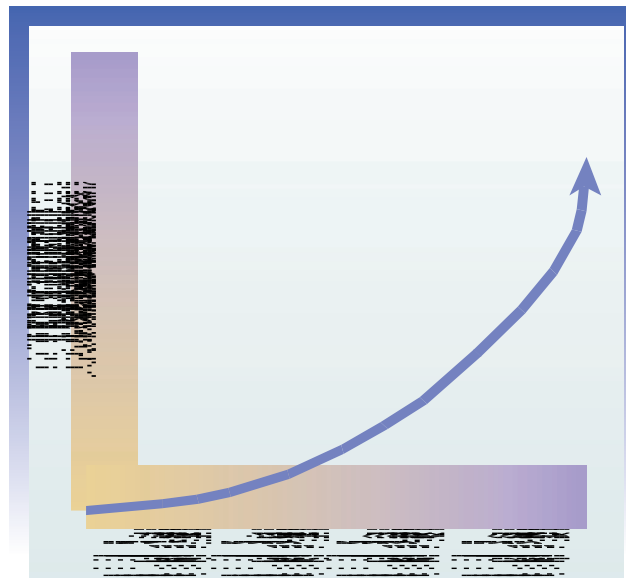
The demand for data seems endless. "You no longer just need to know your customers, but the marketplace and the competitors. And the competitors come from different places," says Terril Retter, a senior analyst at consultant Price Waterhouse in Menlo Park, CA. "The kinds of questions CEOs ask are expanding dramatically." This demand has trickled down, too. Vice presidents, department managers and many others want access to the corporate data stores.

Despite all this activity, even more data is out there. Jim Johnson, chairman of the Standish Group International, a research firm in Dennis, MA, believes business has barely made a dent in collecting data about the customer. He recalls a client, a Las Vegas casino, which knew practically everything about its "high rollers" but almost nothing about the next tier of customers. Businesses can no longer afford this sort of ignorance.

## The Big Picture

A host of other factors also contribute to data growth. There's the technology itself. Relational databases have made information more useful; graphical user interfaces have made it easier to use; and today's end-user tools make analysis efficient and thorough. As the technology gets cheaper and easier to use, even more data is collected and analyzed. Existing applications in some cases are doubling in size, and new apps are needed quickly.

Consider the Federal Express customer service application. It used to be you called up and got one of 50 customer service reps who would report the progress of your package. Then the company extended the application to thousands of drivers with hand-held devices. Next it went to top customers, and there were 50,000 users. Now it's on the Internet, accessible by millions of customers. Each iteration requires more data to make the



To a significant extent, IT itself is responsible for the explosion in data growth. More data is available, more can be done with it, and it is considered to be a critical differentiator between successful organizations and those that fail. This chart, drawn from industry estimates, shows how the base of data has grown and suggests that the phenomenon is only gathering steam.

application work.

What we might call meta-factors also cause data to proliferate. David Flaxman, a partner in the Radnor, PA, office of consultant KPMG Peat Marwick, believes that most industries run in five-year cycles. When the industry moves into a new cycle, it ushers in a whole new set of data requirements. He points to watershed changes in financial services (Wall Street in the 1980s), retailing (thanks to Wal-Mart), telecommunications (the breakup of the Bell system) and commercial banking (in the present day) as examples that demanded new and better data throughout the industry.

Johnson of the Standish Group believes deregulation in certain industries is a key factor. "The more things get deregulated, the more information government requires about the entity. As soon as the airline industry was deregulated, the government wanted more information on things like maintenance and on-time records." A twin cause is litigation, he notes; more information is being archived for much longer periods on a CYA basis.

Not least of all, corporate reengineering or downsizing, and the distributed systems they call into play, are culprits. A downsized company needs to empower the few to do the work once handled by many. IT and better and more data are necessary to do that. Data General's Seban says a driving force in his company's data warehouse strategy was the downsizing of the workforce from more than 18,000 five years ago to 5,000 today. And distributed systems themselves result in small individual or workgroup databases that eventu-

ally must be replicated to a wider audience, producing duplicate databases in various locations on the corporate network.

## Controlling the Uncontrollable

IS professionals must bear the brunt of managing this data rampage: storing data, moving it into relational databases or warehouses where it can be accessed by end users, cleaning and maintaining data, acquiring or building applications for the end users, and training support staff and end users. "How do you control something that is out of control?" asks Katherine Hammer, president and CEO of Evolutionary Technologies in Austin, TX, which develops software tools for data migration and maintenance. "The IS pro needs to accept that he's running a distributed, heterogeneous space and that he's responsible for it." For example, IS has to determine what is the database of record for each different application and control what information goes where and to whom.

This, of course, is easier said than done. The headaches are enormous. Sterling Makishima, HP's data warehouse manager for global customer support, says his biggest problems are maintaining data integrity, assuring the timeliness of the data being demanded by the end user and supporting a worldwide data warehouse infrastructure.

IS people are increasingly turning to some version of the data warehouse to manage this flow. "To improve the look at traditional data, you put it in a data warehouse," says Price Waterhouse's Retter. "You pull it from OLTP, rationalize it and

put it on an OLAP for EIS or DDS.” This also is not simple to do. “Most IT pros are struggling to integrate data across stovepipe legacy apps,” he says.

Whatever the problems for IS, management looks for the advantages. “Data warehouses have a good payoff based on productivity,” says Paul Cabbage, a senior analyst at Dataquest, a research firm in San Jose, CA. “You can get metrics that show how much money you’re going to save by going with the data warehouse concept.”

In a growing number of cases, the business unit—not IS—is proposing and paying for data solutions such as warehouses. In Retter’s view, reengineering changed the perspective from using technology for business functions (vertical) to using it across business processes (horizontal). But vertical data structures don’t match horizontal data needs. “For 30 years the heads of data processing departments made technology sound mystical,” he says. “Now that technology is more important to the health of the enterprise, business has to pay attention.”

This restructuring makes it imperative that IS pros and business users talk to each other. “Business units are driving the data warehouse phenomenon,” says Betancourt of SAS. “IT has to spend a lot of energy transferring the data logic. To do that properly, they have to go out to the business users and ask, ‘How do you want it transformed? What are the business rules?’”

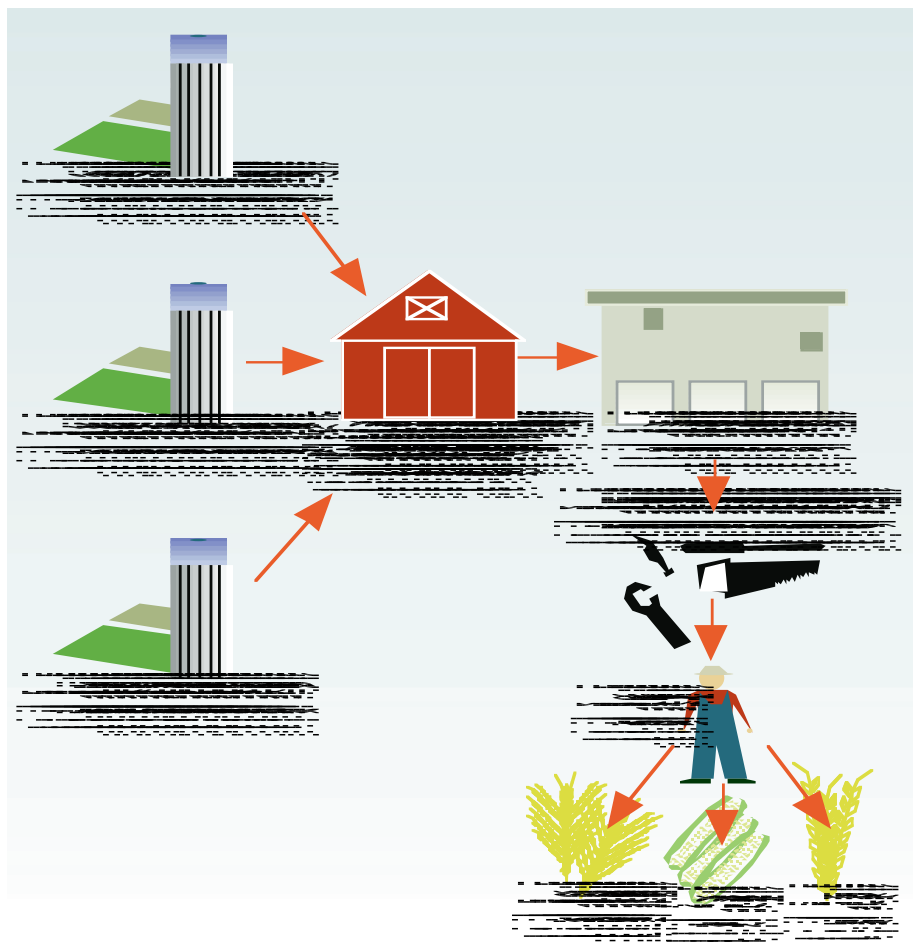
To provide insight into how leaders in technology implementation are solving their data growth and access problems, we present the following somewhat related, yet uniquely defined case studies.

### HP Worldwide Customer Service Organization

#### Asking the Unaskable

Ferguson, the product manager at HP, points out that decision support systems have been around for years, but answering previously unasked questions about customers was a painful process. Because it would take IS two to three months to do a new report, “very few questions were asked,” he says.

Today, thanks to several data warehouses and tools designed for end users, Ferguson and others can get detailed



In an advanced data access architecture, the data warehouse becomes the center. It receives data from many sources, which has been normalized for use, and provides it to users through various tools.

answers quickly. “We can look at our customers more broadly and see the business we’re doing with them,” he says. “We ask a lot more questions.”

That’s what Makishima, the data warehouse manager, intended when his team began building the Worldwide Customer Service Organization warehouse two years ago. Customer support was once merely reactive, responding when the customer called. “Now when the customer’s contract expires, we can go out and inform them of new programs and see if they are interested,” Makishima says. “It gives us a competitive edge.”

Customer service used to be handled by a mainframe at corporate headquarters. With users demanding more detailed information, the system took too long to load data and too long to respond to queries, and it cost too much to support. For those reasons and obvious technological preferences, HP decided to migrate off mainframes.

After investigating several products,

Makishima’s team chose software from Red Brick Systems of Los Gatos, CA, to run on HP 9000 servers. Makishima says IS wanted to get out of the ad hoc reporting role and provide data so users could select what they wanted. Red Brick’s software loads millions of rows of data speedily and supports new data access tools.

There was strong management sponsorship. Marketing and finance lent people full-time to the project. “We had a good idea how users would use it, but when we modeled the Red Brick system we were able to check with them,” Makishima explains. When the team began to build the warehouse, the design called for 20GB. It’s now at 100GB with room to grow to 200GB. “Our users have realized the importance of historical data,” he says.

The biggest challenges appear to be assuring the integrity and timeliness of data, and providing support to users worldwide. “At first we had a handful of users. In a little over a year we’re at more

(CONTINUED ON PAGE 28)

This page  
intentionally left  
blank

(CONTINUED FROM PAGE 26)

than 400 and growing at 18 end users per month," Makishima says. "Originally the end users were just in customer support. When the word got out we had all this consolidated information, other people started using it." The demand is more than Makishima's staff of six can support. HP is adding individual business information managers around the globe, under the budgets of the workgroups they support.

Another problem was that management wanted the job done in three months. "We got it up in about four or five. That was pretty good, because warehouses were taking more than a year," Makishima recalls. "That's one reason we don't always use HP products. We don't have time to reinvent the wheel, so we leverage partnerships with other companies."

### Perkin-Elmer Corp.

#### Out with the Greenbars

In the beginning there were greenbars: those mainframe-generated reports printed on reams of green-and-white-lined paper. The greenbars begat a DOS-based EIS. EIS begat a Unix-based relational

database, which begat four regional data warehouses. Now Perkin-Elmer Corp. is looking at a World Wide Web solution to give one-fourth of its 5,500 employees access to rich data stores.

In 1990, an executive with technical savvy was put in charge of the product group that accounted for half of the \$1 billion revenues at the Wilton, CT-based laboratory instruments manufacturer. He wanted to know what products were ordered where. But all data resided on a mainframe; greenbars came out monthly. Daily greenbars could be cranked out only painfully.

John Stoveken, a consultant, led a team that took a couple of years' worth of data off the mainframe and built a DOS-based EIS. "There were probably 10,000 rows of information," recalls Stoveken, who today is the company's DSS manager. "It was world order information, by country and by product. It started off under 10MB. It was accessed by a small group and grew to 100 users."

The users demanded more. The system grew to 800 product groups. Next, users asked for information on 15,000 to 20,000 active parts. When they asked for the entire

parts list of 130,000, IS saw that it had reached the limits of DOS. About three years ago, a decision was made to move to a data warehouse. A first step was selecting the Oracle RDBMS on a Sun Sparc 1000. "We started to grow the data systems and to look for tools to manage the data we were bringing in," Stoveken says.

During the pilot test, Stoveken discovered CrossTarget, a client/server multidimensional analysis and reporting system from Dimensional Insight of Burlington, MA. The CrossTarget engine uses array indexing to transform data from spreadsheets, SQL databases and legacy systems into multidimensional models. The client, called Diver, lets users view data as they wish, perform analysis and "drill down" into subsets. The models, of 250MB or less, are built from data stored on Oracle and go out over the LAN to be used by anyone who has the client software. This way, Stoveken's 10-person staff keeps control of the data in the warehouses.

U.S. users were happy. "Then people around the world started asking for more detail," Stoveken says. "We said it's time to do data warehousing on a global basis." They chose four HP 9000s running Oracle at manufacturing sites in California, Connecticut, Germany and the United Kingdom. "When you have 40 different order-entry systems around the world, there are coding inconsistencies," Stoveken explains. "So we use the data warehouse as a central point where we can clean data, take out the coding problems and make sure it's a well-defined set."

What's next? With a goal of giving as many as 1,500 employees access, Stoveken is looking at CrossTarget's new Hypertext Markup Language (HTML) version and considering using it over the company Web. "The company runs a WAN that stretches across the U.S. to Germany to Japan. Through our intranet, we can open up some of these larger models to the global users."

### CoreStates Financial Corp.

#### Getting Intimate with Data

If a customer doesn't answer the letter that offered a special interest rate on a new credit card, CoreStates Financial Corp. wants to know why. That level of data analysis requires a staff that is more intimate with the data than are typical IS staffers.

At CoreStates, a Philadelphia-based

## Hard Facts About Data Growth (and How to Deal with Them)

#### Data will grow faster than you expect.

Plan for more than you believe you will have. Include servers, storage hardware and software, and data access tools.



#### Different people want different data or formats.

Examine your existing data in advance. Talk to users. Determine what formats are needed to serve applications.



#### Data warehouses are complex, expensive projects.

Start small, with a prototype. Get management and user buy-in in advance.



#### People are lemmings.

Many more users will want to take advantage of the data warehouse if it succeeds. Build with scalable hardware and software.



#### This market is immature, to say the least.

Avoid vendor lock-in to specific platforms and products. Improved capabilities are on the way.

bank holding company with \$30 billion in assets, the retail credit risk technology unit not only analyzes the data, it manages it, develops its own applications and is building a data warehouse. "Our group is comprised of analytical people, used to dealing with large quantities of data and turning it into competitive advantage," says Jeffrey Oulton, vice president of retail credit risk technology. "They're not reading batch runs and turning them into spreadsheets. They're merging and analyzing data for decision-makers."

Within the 20-person team are people dedicated to building a data warehouse. "They like data and like to code, but they understand business," says Oulton. "They're our liaison to the technology side of the house; their mission is data acquisition and making it available to others."

At the heart of credit risk analysis is scoring. Within the group, one team works on the front end of the business—

determining whether to extend credit. Another concentrates on the back end, traditionally known as collections but increasingly a source of new revenue, including offering bigger lines of credit. These two groups were separate and worked off different sets of data. As a result of a reengineering effort last year, they combined. "The piece that is completely new is the data warehousing," says Oulton.

They're migrating data off a mainframe with the DB2 RDBMS to an IBM RS/6000 server running Oracle. The LAN is Novell NetWare. "The entire corporation is also building its own warehouse. There will be links between them," says Oulton. "We're ahead of the rest of the corporation. The corporate IT guys are not as comfortable with the Unix environment."

Oulton praises the IS side of the house. "They understand that some end users understand the data better than they

do. IT is very supportive of the project. We play on their expertise. Data modeling, loading in the DB2 environment, the decision whether to move to Oracle and how—that's their call."

Using tools and consultants from SAS Institute, Oulton's team has built application suites for three types of users. The EIS is a highly summarized GUI with some drill-down capabilities for about 20 executives. The power-user version satisfies the people in his unit. And a middle-of-the-road suite serves about 200 others who need access to data. "We've been able with one tool to address three levels of user," he says. "That's important from cost justification and support standpoints." ■

*Bill Roberts is a free-lance writer who covers business, technology and management issues. He can be reached at [wcrober@aol.com](mailto:wcrober@aol.com).*