

March 19, 2001

## Golden Data

**By Charles Babcock, Interactive Week**

Data warehouses, data mining systems and business intelligence applications stand in readiness to serve e-commerce, but they carry one Old Economy characteristic that gets in the way: The information they're working with is several days — or, more likely, weeks — behind events.

In an online environment, business intelligence users say, that simply won't do, and they want to increase the speed at which information is updated.

"A successful supply chain needs to be tied to the best patient outcomes," says Chris Stewart, director of data warehouse services for the Health Care Informatics division at national hospital alliance Premier. By mining patient data, 1,700 health-care providers in the alliance can discover which drugs and other supplies work best. "It's our job to provide the best hospital services possible. The data we collect is crucial to that," Stewart says.

And the data that Premier's Informatics division collects is more up-to-date than it might be otherwise because of a feature called "versioning" in its Red Brick data warehouse system; Red Brick is a division of Informix Software.

Versioning allows multiple queries of the same data. It allows the creation of

database tables that can be queried by the purchasing agent at one hospital, without freezing out different queries that want to use the same data. Unlike many data warehouse systems, the patient data that makes up those tables can be updated in the background, without disturbing the first purchasing agent's queries. Queries coming slightly behind the purchasing agent's will still get access to the same data and, in some cases, it will have been refreshed with the latest information — all without disturbing the initial query.

Versioning also allows the most recent data to be continually added to the data warehouse and appear in the latest query, says Fred Ho, executive director of engineering at Decision Server, Red Brick's data warehouse.

That move was a major step forward for a user such as Premier, which has 750 gigabytes of data and many queries. New information can be loaded into the data warehouse and subjected to a query "in a matter of minutes now," Stewart says.

The advent of multitudes of business intelligence application users has posed its own problems, says Eric Miles, senior vice president at Sybase's Business Intelligence division. For example, when Telstra, a cellular phone service provider in Australia, found its system overwhelmed during the Olympics last summer, it

programmed its antenna pointing system based on previously discovered calling patterns mined from the data warehouse. For example, Telstra found traffic mushroomed at the end of the gymnastics events as crowds of Chinese onlookers left the stadium to call in results, and adjusted its system accordingly, Miles says.

Sybase uses IQ Multiplex, its variation on indexing data, to speed data access and return results from what the company calls its "portal-ready" data warehouse system. IQ Multiplex's indexing and rapid retrieval was behind the system ranked "as the largest data warehouse in the world on an NT platform," says Richard Winter, president of Winter, which annually ranks the largest database systems in the world.

To speed the use of Web site data, SPSS, a Chicago data mining software vendor, brought out its Clementine 6.0 workbench in December. Customers use the workbench to build online data mining applications, and SPSS has added Clementine Application Translator templates to the workbench, giving Web site developers "80 [percent] to 90 percent" of the framework of an application, says Colin Shearer, vice president of data mining at SPSS. By using the templates, developers can rapidly construct profile engines and other applications that respond to visitors on a site, using the data about them that's in the data warehouse, Shearer says.

In a poll of 287 Web developers, Clementine was the tool of choice for 21 percent, followed by the SAS Institute's Enterprise Miner with 17 percent and the Berlin-based Humboldt University's Web Utilization Miner with 16 percent, according to data mining newsletter and Web site KDnuggets. WUM is designed to mine the data in Web server logs for user behavior patterns.

In similar fashion, Sagent, a supplier of data mining tools, announced in February a move into analytic applications such as Web Analysis Solutions, which pulls together clickstream, U.S. Census Bureau and other demographic data and business data, says Ben Barnes, president of Sagent.

The next step, Barnes says, is to link these applications to an upcoming Event Server, which will respond to predictive models built from the data warehouse and respond to defined events. An airline Event Server, for example, might "see a flight not filling up at the pace it should and open up more discount seats," Barnes says.

Usama Fayyad, chief executive of online data warehouse service digiMine, says Web site developers are torn between using clickstream analysis and a preconstructed profile in deciding how to respond to visitors with special offers.

DigiMine will add a site's customer data to its warehouse and build categories of visitors with it, turning to those categories when its best information on a visitor is a clickstream that fits into one of the categories. In other cases, the visitor may have entered an ID that, combined with the visitor's history and clickstream, provides another set of information on how to respond.

In either case, Fayyad says, digiMine monitors the site and reports how many visitors are converting into buyers or whether frequent visitors are decreasing their visits. The result, he says, is that a site manager "can respond right away, instead of a week or two going by without your noticing something went wrong."