SUBSCRIBE TODAY!

Front Page Articles

November 2003:

- More Expansive, Less Expensive BI By Joe McKendrick
- Inside the Sloan Digital Sky Survey By Walt Jordan
- Supporting a Culture Shift By Billy Rosario

Front Page Archives

- October 2003
- September 2003
- August 2003
- July 2003
- June 2003
- May 2003
- April 2003
- March 2003
- February 2003
- January 2003
- December 2002
- November 2002

FRONT PAGE

November 2003

More Expansive, Less Expensive BI

By Joe McKendrick

The business intelligence market has long been fractured, forcing database managers to cobble together solutions from multiple vendors to include components such as data warehouses, front-end tools, and ETL tools. Over the past year, however, a dramatic consolidation has been taking place within the BI sector, heralding the potential emergence of a more expansive-- and potentially less expensive -- generation of BI and analytic tools. "Business intelligence -- or enterprise reporting -- used to be a niche, but now it's a full-scale category," Gerald Cohen, president and CEO of Information Builders Inc., told DBTA. In turn, the growth of BI has fostered many subcategories, he continued. "There are vendors that emphasize business analytics and OLAP. There are others that specialize in data mining. There's the do-it-yourself reporting vendors, and the enterprise reporting providers."

The consolidation is taking place at two levels. Across the industry, there has been a spate of acquisitions as vendors move to offer the essential ingredients of BI under one roof. Within the products, themselves, there is a movement to consolidate information from across all areas of the enterprise into a single dashboard. "Companies can't afford to have multiple versions of the truth," said Neil Mendelson, vice president of business intelligence products for Oracle Corp. "That becomes very costly from a time and cost point of view. A few years ago, there was a move toward data marts, which were successful in moving a lot of data out to them. Unfortunately, they got the data in these little bitty data marts, and in Excel spreadsheets. Everybody fractured their information flow to the extent that no one had the right answer --or at least a single answer. Now we're seeing a move back to consolidation."

Over the summer, Hyperion announced its acquisition of struggling BI software provider Brio Software. Right before the Hyperion announcement, Business Objects announced its purchase of Crystal Decisions, a well-known analysis and reporting tool provider. At the end of 2002, Cognos acquired Adaytum, an enterprise planning tools vendor. Other deals shaping the industry include Lawson Software's acquisition of Closedloop Solutions, which sells budgeting and financial forecasting software. "They're filling in the missing pieces," said Chris Silbernagel, managing principal at Exigent Partners, a BI consulting firm. "The large analytical BI vendors are adding report-writing capabilities. The ETL vendors are adding analytics. And the database vendors are adding both."

For Hyperion, the Brio acquisition is "allowing us to fill in some gaps with our tool offering," John O'Rourke, senior director of product marketing at Hyperion, told DBTA. "We weren't as strong as we wanted to be in terms of offering tools that can query directly against relational databases, transactional systems, and data warehouses."

For Business Objects, acquiring Crystal Decisions provides an entrée into the developer marketplace. "Business Objects has had a much more direct enterprise sales model, while Crystal has a fantastic developer network," said Lance Walker, director of product marketing at Business Objects. Cognos unveiled ReportNet, an enterprise-reporting tool that integrates with Cognos Enterprise BI solution, which incorporates scorecards, dashboards, OLAP analysis, event detection and alerting, and data integration. "Companies have invested enormous amounts of money in their data transaction-based systems," said Michael Branchaud, director of BI product marketing for Cognos. "This decade is about taking information from disparate systems and delivering that information to business people at all levels."

BI tools vendors are building out their suites to compete against the enterprise application vendors--Oracle, SAP, PeopleSoft, and Siebel--hungrily eyeing this market. The Microsoft juggernaut is beginning to move into the low end of the BI space. The BI segment has performed far better than most software companies through the IT slowdown over the past three years. That's because BI has a relatively low cost of entry and leverages existing IT resources. However, if end-users need to do a great deal of customization, integration, or are unable to reach key data sources, the value of BI may be lost. The proliferation of smaller BI tool vendors was fueled by the fact that large enterprises required products from at least four or five BI vendors to fully cover all the separate aspects of BI -- such as relational databases, multi-dimensional databases, OLAP, data mining, ETL, reporting, and data

warehouses/marts. Said Larry Barnes, BI practice director for Avanade, "Historically, customers would purchase separate tools for best-of-breed capabilities for each area of BI. However, this approach has only driven up acquisition costs and also created the need for specialized resources for both development and ongoing operations."

The need to integrate and consolidate is being felt. "Small vendors created market niches and that resulted in a situation that became untenable to businesses, which had to manually integrate many niche products," agreed Eleanor Taylor, manager of BI strategy at SAS. "Customers now demand integrated suites of products that work together by design. Niche vendors are faced with having to scramble to acquire or develop the capabilities they don't have, or risk being acquired by a competitor."

Another powerful force that is reshaping the industry--and also driving vendors into higher-end suite offerings - is Microsoft's aggressive moves into the BI market. "Microsoft's Analysis Services provides affordable BI," said Avanade's Barnes. "It comes bundled with the SQL Server product line, which is already the low-cost provider in the relational market. That means customers no longer have to spend six figures or more for scalable BI. In addition, the tight integration with the relational database and the inclusion of an ETL tool--DTS--into SQL Server started the trend towards integrated BI suites. IBM and Oracle had to fill out their product lines to match what was available in SQL Server."

"Yukon [the next release of SQL Server, scheduled for the second half of 2004] takes this to an even higher level," Barnes continued. "We've been working with the beta over the last few months. DTS has grown up. The Universal Data Model provides a metadata abstraction layer that allows one common definition over both relational and multi-dimensional models."

Some leading end-user companies are beginning to consolidate their own BI development efforts. A push for integration is on at Nielsen Media Research, where ongoing analytical requirements -- both internal, and extended to customers -- require more integrated capabilities, according to Kamal Nasser, vice president of IT strategy for Nielsen Media Research. Nasser told DBTA that Nielsen employs just about all the major BI tools to access data residing within a Sybase IQ data warehouse on the back end. Now, the large services company is turning to Microsoft Analysis Services for most of its front-end BI requirements. "We realized we can't have every single application tapping directly into the data warehouse," said Nasser. "So we have to create subject-area data marts, or cubes. We're relying on the new-generation .NET technologies, like Analysis Services to accomplish that."

Many BI vendors are repositioning their offerings as business performance management tools. BPM focuses on data coming in from across the organization, versus a specific area. "BPM applications are used company-wide and fairly deep," said Craig Schiff, founder of Hyperion and now CEO of BPM Partners. "They cut across multiple departments, and they reach pretty deep into those departments. BPM is the application of many of the components of BI, comprised of a multidimensional database, an OLAP cube, report and query capabilities, ETL, the ability to move the data back and forth, analytic applications such as planning, executive dashboards or score cards, and portals to deliver all of this information. Companies have been trying to assemble BI components, themselves. They're finding it's quite difficult to do that and make it all work together."

BPM is the result of three major markets colliding, said Aaron Zornes, vice president with Meta Group, financial application vendors, like Hyperion, mega-application vendors like Oracle, PeopleSoft, and SAP and the business intelligence vendors such as Cognos, Business Objects, and Informatica.
BPM will also function to provide real-time snapshots of the business. "We see a trend towards real-time visibility and the ability to take action on events that happen," said Peter Galvin, vice president of marketing for Celequest Corp. Visibility into actionable events can lead to better products, reduced risk and improved customer service."

For more articles by Joe McKendrick, click here

back to top

Inside the Sloan Digital Sky Survey

By Walt Jordan

Making maps is one of the most traditional forms of recording scientific knowledge. In the past, people made maps of the worlds they explored on foot or by boat. These days, scientists make maps of the worlds they observe through other methods--the microscope and telescope.

The Sloan Digital Sky Survey (SDSS) is an ambitious project to systematically map one quarter of the sky, including all of the northern sky and a slice of the southern sky. The goal is to produce detailed digitized images and to determine the position and absolute brightness of more than 100 million celestial objects. A joint project of The University of Chicago, Fermilab, the Institute for Advanced Study, the Japan Participation Group, The Johns Hopkins University, the Max-Planck-Institute for Astronomy, New Mexico State University, Princeton University, the United States Naval Observatory, and the

University of Washington, the project team also hopes to provide a three-dimensional picture of the distribution of matter to the edge of the visible universe. Apache Point Observatory, site of the SDSS telescopes, is operated by the Astrophysical Research Consortium (ARC).

SDSS marks a shift in the traditional methodology in astronomy, which has generally focused on studying one object at a time using general-purpose telescopes. "Instead of studying a few objects closely, we are surveying a huge swath of the sky and then using data mining, hopefully, to find interesting things," said Jan Vandenberg, the self-described systems and database nerd for the project at The Johns Hopkins University, where the project's data archive and the data catalog server has been developed and maintained. The data catalog server can be viewed as the data mining interface to the data.

More than 200 scientists from participating and non-participating institutions are associated with the project. Data is captured by the SDSS observatory in at Apache Point Observatory in the Sacramento Mountains of New Mexico and then processed at the Fermi National Accelerator Laboratory in Illinois. It is the passed to The Johns Hopkins University for access by scientists and then the public.

"In addition to being interesting from the point of view of astronomy, the project is interesting from a database and computing point of view," Vandenberg noted. The first data release last summer contained more than 80 million unique objects and each object in the survey has 400 data points associated with it, including information about its shape, color and what kind of object it is. In addition, the data catalog has more than 160,000 spectrographic objects, each with hundreds of associated data points as well. In total, the first release of the data had almost 3 TB of information. Data Release 2, which is scheduled for this month, will have 50 percent more data. Eventually the project could generate around 40 TB of data, made up of 25 TB of source data and another 13 TB of processed data.

Both the amount and nature of the data raised significant issues. How should celestial objects be stored in a database? How should information be indexed to ensure speedy retrieval? How could spatial relationships between objects be represented and maintained? For examples, researchers may want to find the objects clustered around a certain point in the sky. Finally, could the data be made available to the public via the Internet in a way that could be fun and useful?

Five years ago, when Vandenberg first joined the project, the team was using an object-oriented database for the catalog server. "On paper, an object-oriented system sounded like just what the doctor ordered," Vandenberg recalled. With millions of objects in the sky that could be linked to each other, an object-oriented approach seemed right.

But the object technology the team selected didn't meet their needs. Although the database itself is not terribly large for a data warehouse, users could easily and routinely write queries that would call on the entire database. Consequently, speed and disk I/O were significant issues. Moreover, the system is intended to be interactive, though, Vandenberg noted, it has still not yet reached the level of interactivity the scientists want.

The object-oriented database technology was simply not fast enough, Vandenberg said, nor could it be optimized in a suitable way. And there were manageability issues; particularly concerned with integrating some of the original code researchers were developing. "We wanted hundreds of megabytes per second of performance," he said. "We needed direct disk I/O."

As team members under the leadership of Alexander Szalay faced the problems with the object-oriented data approach, they also began to explore how they could address another goal of the project--to make the data available to the public in a timely fashion.

Public access to the data is hosted by a project called SkyServer (skyserver.sdss.org/dr1/en/). "It gives users full access to the data packaged in an easy to use interface," Vandenberg said. "You can navigate the database and it is like having your own telescope." The images are pre-computed jpegs. But all of the data comes live from the database. There are 14 million objects in the early data release of SkyServer. In March, the site generated more than one million hits.

With these two issues in mind, the team began to explore options. In that process, they began to work with Jim Grey, a database expert at Microsoft. As the relationship grew, the choice to use SQL Server seemed natural.

Initially, the Hopkins group developed a relational database to run along side the object-oriented database. "The astronomers were using the object-oriented database and the public outreach was on SQL Server," said Vandenberg. It quickly became clear that it was quicker, easier, and less expensive to just use SQL Server. Porting the database from the object-oriented platform to SQL relational model was not trivial. Microsoft has published a technical article describing the process at ftp.research.microsoft.com/pub/tr/tr-2001-104.doc.

As an added benefit of moving to SQL Server, although the astronomy community almost exclusively uses Sun Sparc workstations running Solaris, the catalog database is hosted on low cost x86 Intel platforms. About three years ago, Vandenberg said, the x86 emerged as a fast, reliable server, able to compete with Alpha and Solaris servers.

Storage has also been an issue. "It has always been hard to have enough spinning disks on a research budget," Vandenberg observed. That problem has been resolved by the emergence of ATA

Raid storage. The data processing operation at Fermilab has 40 TB of ATA Raid storage. And there is a 3 PB tape archive.

In addition to creating a relatively low cost, high performance infrastructure for the data catalog, the Hopkins team has developed some innovative database techniques. For example, to enable scientists to quickly analyze clusters of objects, the team developed a spatial indexing scheme, called the Hierarchical Triangular Mesh, directly in the stored procedures. "It is an interesting and efficient way to create a spatial index for a spherical data set," said Vandenberg.

In it, the spherical space is divided into a series of triangles. Penetrating several layers provides a high degree of granularity for the data. Combined with innovative algorithms, the indexing scheme accelerates queries, allowing people to explore huge amounts of data quickly. "We don't want to use the database server just as a repository for people to get data out. We are trying to do as much of the work in the database. So you will get an answer that you find interesting right off the bat," Vandenberg said.

The project's payoff has been impressive. In the early days of the project, SDSS scientists identified several instances of a new type of astronomical object, so-called brown dwarfs that are smaller than stars, but larger than planets. Until then, only one such object had ever been detected in the universe.

Since then, there has been a steady stream of news including the observation of the most distant object ever, new quasars and eight new galaxies that may be the youngest in the universe.

While the data catalog server runs primarily on 32-bit platforms, Vandenberg and his team have experimented with an Itanium platform and 64-bit SQL Server. "We were able to do all of the analysis for one problem in RAM on the Itanium," he noted. "It has been a straightforward process to convert simple Fortran programs to the Itanium and move obscene quantities of data in RAM."

In the long run, he believes moving to 64-bit systems will result in a "huge win." It will enable a new class of experiments to be conducted, he said, at a reasonable cost.

back to top

Supporting a Culture Shift

By Billy Rosario

Measure, analyze, control, automate and improve process-related operations. Emerson Process Management supplies products, services and solutions worldwide to achieve those tasks. The company is a subsidiary of St. Louis-based Emerson, a $13.8 billion global technology and engineering company, which has long had a product-oriented, engineering-based culture.

During the go-go years of the 1990s, Emerson Process Management went on an aggressive acquisition spree to build its business. As it grew, it found itself facing stiff competition across many of its lines of business, particularly in the electronic controls market sector, in which a major division operated. Despite significant capital investments, many of the company's products had become commodities. Moreover, the worldwide market in the use of electrical energy had flattened out, reducing Emerson's ability to maintain and raise its prices. Despite gains in market share, the company's performance suffered.

And there were other challenges. The company's growth had left it with a complex business organization encompassing more than 160 different units, each with its own information infrastructure. Further complicating matters, Emerson Process Management sells its products through both direct and indirect sales channels. Of its top five divisions, two have a direct sales force and three use independent sales representatives. And despite the fact that Emerson had more than 1.5 million customer records scattered through its systems, a significant amount of customer contact information was never captured in the corporate databases.

As a result, according to Nancy Rybeck, Emerson Process Management's customer data warehouse strategy architect, sales personnel in one division could not know if another division was also doing business with a specific company. Consequently, the company could not identify its most valuable customers, nor could it easily bring to bear all the resources of its relevant divisions to provide more comprehensive solutions for its customers. In short, the diffusion of information among its divisions made it difficult for Emerson to broaden and deepen its relationship with its customers.

As the overall market slowed and prices weakened, company CEO David Farr decided to shift the company's orientation from a product-centric approach, to a customer-centric, solutions-oriented approach. Emerson would improve its growth rate by expanding its relationships with its customers. In addition to selling products, the company would offer pre-sales consulting and post-sales training. It would offer maintenance. And would aggressively look for cross-selling opportunities.

The challenge was to create an information infrastructure that could support the new approach. According to Rybeck, the company had built a data warehouse. "But it was hard to use," she said, and

never met the company's expectations.

So the company set out to institute processes that would allow it to capture all customer transactions regardless of where they occurred and push this information to people in the field. Sales, marketing and service personnel could then use this information to identify additional opportunities for Emerson within each customer.

The new system was based on an in-house e-business application dubbed the Project Account Tracking Tool (PATT) and a data warehouse built on an Oracle database. An essential step in the process was a rigorous program for data cleansing and de-duping of the customer records. For that, Emerson used Group 1 Software's enterprise data quality solution, DataSight 3.0, to standardize, match and consolidate customer data. Rybeck had used Group 1 solutions earlier and felt the company had the best global support. Finally, Dun and Bradstreet DUNS numbers are appended to each record, to serve as a cross-reference number for global companies. The initial implementation took around 11 months to complete and is now updated every six to eight weeks.

The data cleansing and de-duping process had several interesting challenges. Not only is information entered around the world daily in different languages and multiple formats, but commonly the ordering company and the company to which the products are delivered are not the same. Moreover, since Emerson supplies products to new plants just being built, often there is not a specific address or Dun and Bradstreet number to append.

Investing in the data quality initiative has paid off big dividends, Rybeck said. Showing sales representatives the relationships each division of the company has with customers opened up new sales opportunities. "It was like fruit falling at their feet," she said.

The company could finally accurately identify its very best customers. "We could show them that instead of there being two companies among our top 50 customers, it was really one company in our top 5," she said. Over time, Emerson was able to increase its share of purchases per customer at a double-digit rate, reduce its service costs and record better levels of customer satisfaction. It also has won recognition for its strong relationships with its customers.

There were several factors that contributed to the project's success, Rybeck said. Perhaps the key, however, was to be able to provide real benefits to the people in the field. "The sales people had to see it was worth it," she said. "It is not just a reporting tool. You have to drive the information to the people who can use it." In fact, the system generates only a few management reports, and those are often of interest to the field teams, as well.

And, Rybeck said, she learned several lessons from the project. Data cleansing should not be done at the operational database level, she argued. "The operational systems are too big. You can't force it into the operational system," she said.

Secondly, the project has to return real value to make the project worthwhile. The field personnel, for example, have to be able to trust the information they receive. Finally, she noted, "Organizational process changes have to occur at many different levels."

At the bottom line, the CRM and data warehouse project has enabled Emerson Process Management to shift the way it does business. "We are no longer just order takers for commodities," she said. "We sell solutions and that is how you drive value."

back to top

Database Trends & Applications, c/o Unisphere Media, 229 Main Street, Chatham, NJ 07928 ■ Tel: 973-665-1120 ■ Fax: 973-665-1124 ■ e-mail:info@dbta.com