101

BEST PRACTICES IN BUSINESS INTELLIGENCE AND DATA WAREHOUSING

What Works

Powerful Case Studies, Lessons Learned, and Q&A from Leading Technology Providers and Experts Worldwide

in Sponsorship with TDWI

Volume 21

A 101communications Publication

tdwi
THE DATA WAREHOUSING INSTITUTE

LETTER FROM THE DIRECTOR

A wave of acquisitions, a host of start-ups, and significant innovations in hardware—all combine to make business intelligence (BI) the hottest ticket in enterprise technology today. Keeping abreast of so many developments can prove a daunting challenge for even the savviest executives; and falling behind the curve is not an option.

With all that in mind, TDWI proudly presents volume 21 of *What Works: Best Practices in Business Intelligence and Data Warehousing*. This compendium of content focuses specifically on BI and data warehousing (DW), providing an information resource for better understanding the tools, technologies, and methods that are central to this growing industry.

The goal of *What Works* is to provide a snapshot of the most innovative BI and DW implementations available today. The case studies included in this volume demonstrate the power of BI for industries ranging from publishing to healthcare, manufacturing to entertainment.

In addition, *What Works* includes articles from leading experts in the services, software, and hardware vendor communities. These "Lessons" feature perspectives about how to build successful BI environments—including such topics as master data management, customer data integration, performance ETL, and data mart consolidation, to name a few.

Our Q&A section provides answers from these experts to the questions they hear most often, complemented by insight from an independent consultant. You'll also find excerpts from TDWI's most recent research reports: *Data Integration: Using ETL, EAI, and EII Tools to Create an Integrated Enterprise* and *Taking Data Quality to the Enterprise through Data Governance*.

The feature article, "Business Intelligence 2006—Only the Beginning!" comes from Wayne Eckerson, TDWI's director of research and services. Here, Eckerson takes a look at the evolution of BI, which he describes as a rich and ever-expanding field. Breaking BI down into five dimensions, he discusses past developments, as well as technology innovations to watch going forward.

We hope you enjoy this collection of case studies, best practices, and expert insight, and we look forward to your comments. If there is anything we can do to make this publication more valuable to you, please let us know. And please join me in thanking the companies who have shared their stories and successes, their technology insights, and the lessons they have learned.

Denelle Hanlon
Editorial Director, *What Works*
TDWI
dhanlon@tdwi.org

What Works
BEST PRACTICES IN BUSINESS INTELLIGENCE AND DATA WAREHOUSING

tdwi
THE DATA WAREHOUSING INSTITUTE

5200 Southcenter Blvd., Suite 250
Seattle, WA 98188
206.246.5059 Fax: 206.246.5952
info@tdwi.org, www.tdwi.org

Editorial Director	Denelle Hanlon
Production Manager	Marie McFarland
Production Editor	Jennifer Agee
General Manager	Meighan Berberich
Director of Research	Wayne Eckerson
Director of Marketing	Michelle Edwards
Art Director	Deirdre Hoffman

101communications

Director of Print Production	Mary Ann Paniccia
Production Manager	Tanya deSomov
Chairman	Curt Hessler
President & CEO	Jeffrey S. Klein
Exec. Vice President	Gordon Haight
Sr. Vice President & CFO	Stuart K. Coppens
Sr. Vice President & General Counsel	Sheryl L. Katz
Sr. Vice President, Human Resources	Michael J. Valenti
Vice President, Business Technologies Group & New Bus. Development	Ellen Hobbs

Sponsorship Opportunities: Denelle Hanlon, 5200 Southcenter Blvd., Ste. 250, Seattle, WA 98188; 206.246.5059 ext. 102; dhanlon@tdwi.org.

Reprints: For editorial and advertising reprints of 100 copies or more, and digital (Web-based) reprints, contact PARS International at 101reprints@parsintl.com or 212.221.9595.

List Rentals: For e-mail and postal list rentals, contact Worldata: 3000 N. Military Trail, Boca Raton, FL 33431-6375; 800.331.8102 or 561.393.8200; 101@worldata.com; www.worldata.com

Corporate Headquarters: 9121 Oakdale Ave. Suite 101, Chatsworth, CA 91311; www.101com.com.

TDWI is a division of 101communications LLC based in Chatsworth, California.

What Works

BEST PRACTICES IN BUSINESS INTELLIGENCE AND DATA WAREHOUSING

TABLE OF CONTENTS

Case Studies by Industry

CASE STUDIES

LESSONS FROM THE EXPERTS

RESOURCES FOR BI & DW PROFESSIONALS

Business Intelligence 2006—

Only the Beginning!

By Wayne Eckerson, Director of Research and Services, TDWI

The Evolution of BI

Business intelligence (BI) is supposed to be a mature industry. It "officially" started in the early 1990s with the emergence of Windows-based query and reporting tools offered by a few start-up vendors such as Business Objects, which now happens to be a $1 billion–plus Goliath in the industry. After 15 years or so—an eternity in software markets—you would expect everything to be sorted out, with a few major players dominating the landscape. Yet, that hasn't happened.

Technology Innovations. I am continually amazed at the number of start-up companies in the BI space that offer new and innovative products. While there has been tremendous consolidation spearheaded by a few leading players, there is plenty of room for nimble start-ups to nibble around the edges of this fast-moving market. In fact, many start-ups are doing more than nibbling—they are gouging rather unseemly holes in the underbellies of the dominant contenders. They are doing this by leveraging:

- New Web application development techniques (such as asynchronous Javascript and XML [AJAX] and Flash technologies).
- In-memory processing that delivers faster processing and 64-bit chips that offer larger memory addresses to cache and process data on the fly.
- Appliances, Web services, and open source software that undermine current pricing structures and open the door to new ways of packaging and delivering content and BI capabilities.
- Google-style user interfaces to simplify access to both structured and unstructured data.
- Service-oriented architectures that create plug-and-play environments that can be quickly enhanced and deployed.
- Dashboard and scorecard applications that conform to the way users work rather than force users to conform to the way the tools work.
- Real-time processing via a bevy of data integration suites that integrate data adapters, ETL, enterprise system buses, federated queries, trickle feeds, and changed data capture.
- Search indexing, text mining, and natural language processing to ferret out the meaning of unstructured data and blend it with structured data.
- Advanced data visualization and analysis techniques that expose patterns and trends in numeric data that are otherwise difficult to detect.
- More flexible and better integrated "what-if' modeling to support planning, forecasting, and simulation applications.
- Strategic management methodologies such as balanced scorecarding and business performance management.
- Extranet security and large-bandwidth Internet pipes that allow users to invert their data warehouses, and BI tools to make information and insights available to customers and suppliers on a massive scale.

Given this raft of new technologies and approaches, the BI industry still has plenty of room left to grow. The challenge for the dominant players is to innovate as quickly as the start-ups—or

acquire these would-be competitors and their innovations—before they erode the established order and the premium pricing that accompanies it.

Five Major Dimensions

There are five major dimensions to BI: reporting, analysis, planning, monitoring, and advanced analytics. See Figure 1. (And many may argue there are additional dimensions.) Each of these dimensions is continuing to evolve and expand as organizations seek to deliver greater insights to more users in a timely way. The following sections describe the evolution of these dimensions and what new developments to expect in each during the next few years.

1. Reporting Dimension

Reporting is the oldest and most dominant of the BI dimensions. In the beginning, organizations simply delivered highly formatted reports to users. The reports were predefined, static, and paper-based. They reported on past activity and events, informing users of what had already happened in a relatively comprehensive way.

Linked and parameterized reports. With the advent of the Web, reporting moved online and became more interactive and dynamic. Developers created linked reports that made it easier for users to navigate from one report to another or from a master view to a detailed view in the same or different reports. Developers also created parameterized reports that let users filter a predefined report and data set using pick lists or prompts. Elated by this newfound flexibility, many users now equate parameterized reporting with ad hoc querying, which it isn't. (Unlike parameterized reports, ad hoc query tools place few constraints on what dimensions, metrics, and data users can view.)

Dynamic push-based reporting. In the near future, Web-based reports will reach a new plateau of interactivity when developers leverage rich Internet application (RIA) techniques, such as AJAX and Flash technologies, to enhance users' reporting experience. (Google was one of the first to popularize the use of AJAX as a development technique in its Google Earth and Google Maps Web applications.)

The Five Dimensions of BI

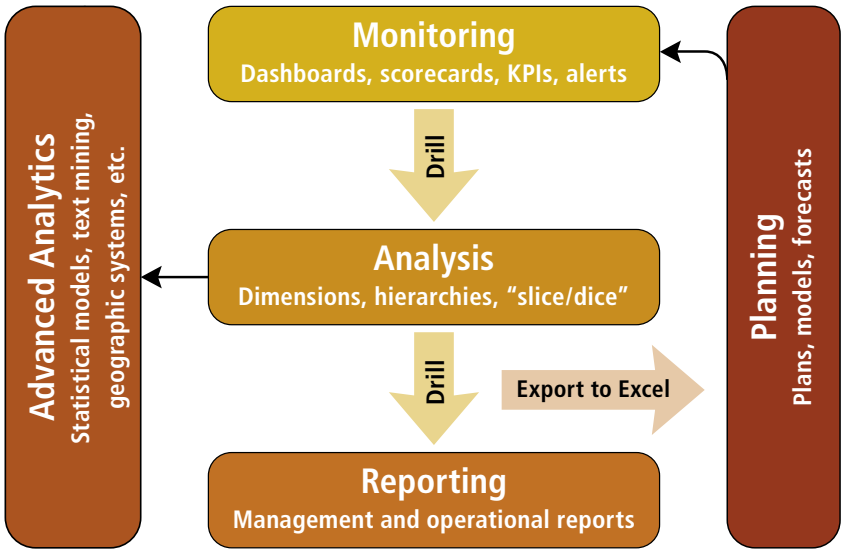


Figure 1. *During the 1990s, BI tools evolved from reporting to analysis tools, which have recently been joined by planning and monitoring tools and techniques. Advanced analytics represents a substantial extension to the analysis layer through technologies that provide broader, deeper, and better analysis.*

While RIA will infiltrate all types of Web-based BI applications, its greatest impact will be on extranet reporting. Instead of distributing 401(K) or payroll statements via PDF, as many financial services firms and providers do today using production reporting tools, companies will deliver secure RIA-based reports that let users interact with data in new and exciting ways. Using the bursting (i.e., broadcasting) capabilities of BI tools, companies will e-mail users customized views of their data—little sandboxes of data and BI functionality—that they can slice, dice, reformat, and revisualize however they want. TDWI calls the intersection of RIA and extranet reporting *dynamic push-based reporting.*

2. Analysis Dimension

In the early 1990s, reporting became passé. Users who bristled under the restrictions of inflexible and non-interactive report formats clamored for more dynamic analytical tools that would allow them to explore and analyze data in a relatively unfettered fashion. Consequently, vendors delivered ad hoc query and OLAP tools that let users "slice and dice" data to their hearts' content. Subsequently, OLAP became really hot and the analysis of dimensional data emerged as the second major dimension of BI.

However, two things happened to cause a major crisis of faith among BI tools vendors whose flagship products were oriented towards analysis. First, it turned out that only a small minority of users really had much interest in using ad hoc query and OLAP tools. Most users found these tools much too hard to use or navigate without getting lost in the data or the tools. "Slice and dice" turned into "lost and frustrated" and created a host of BI shelfware.

Second, most BI vendors abandoned reporting, but then—almost too late—rediscovered that most users actually prefer viewing reports to exploring data in an unfettered manner. In the past several years, leading BI vendors have scrambled to cover their reporting flanks by developing or acquiring reporting tools. That's why Business Objects acquired Crystal Software, Cognos developed ReportNet,

and MicroStrategy and Microsoft both released new reporting products. With both reporting and analysis tools in hand, users would have a complete set of BI tools—a BI suite—or so vendors thought.

In-memory OLAP. Today, a slew of start-up vendors are offering next-generation OLAP technology. These tools avoid the classic problems associated with OLAP cube technology (cubes arrange data along predefined business dimensions for easy access and analysis): limited data sets, a minimal number of attributes, long precalculation times, client tools with limited graphical expressiveness, and headaches associated with managing burgeoning cube farms. Second-generation OLAP products use in-memory-based processing to circumvent performance and management problems, advanced visualization and dashboard displays to enrich user experience, and more integrated modeling (versus exporting to Excel) for planning and forecasting. The analysis dimension is one of the most active areas of BI innovation today.

3. Planning Dimension

By the late 1990s, BI vendors gradually discovered that most users were using analysis and reporting tools as glorified extract programs to dump data into Excel for planning, forecasting, and modeling, or into PowerPoint to create executive briefing books and presentations. These Excel users often issued runaway queries against the BI server so they could get a sufficiently large data set against which to do their analysis and modeling. This bogged down query performance for the few users who were trying to slice/dice data online, causing many to stop using the BI tools altogether. Subsequently, many BI vendors began to rally against the "improper" use of MS Office tools.

MS Office integration. Today, rather than fight Excel and MS Office, many BI vendors have embraced them. They've supplemented their Excel export functions with SmartClient add-ins (i.e., downloadable and dynamically updated plug-ins to MS Office applications) that turn Excel, PowerPoint, and the rest of the MS Office suite into rich clients of BI application servers. Now, users can view and interact with predefined BI reports that are rendered in Excel (with varying degrees of fidelity) and dynamically refreshed when new data arrives. In essence, BI vendors have turned Excel into a full-fledged BI client.

Business performance management. More importantly, leading BI vendors have embraced planning, forecasting, and modeling activities as an additional dimension of BI. More accurately, vendors see business performance management (BPM) as a new and powerful management discipline that uses BI technology as a key enabler and allows vendors to sell products and services at a more strategic level—essentially, they can sell solutions to businesspeople instead of tools to technologists. BPM brings planning, budgeting, and forecasting out of the Excel closet and wraps it in a management methodology that aligns users and processes to strategic objectives and goals and leverages BI capabilities to measure and monitor the resulting performance.

4. Monitoring Dimension

In the early part of this decade (2000–2005), many BI vendors recognized

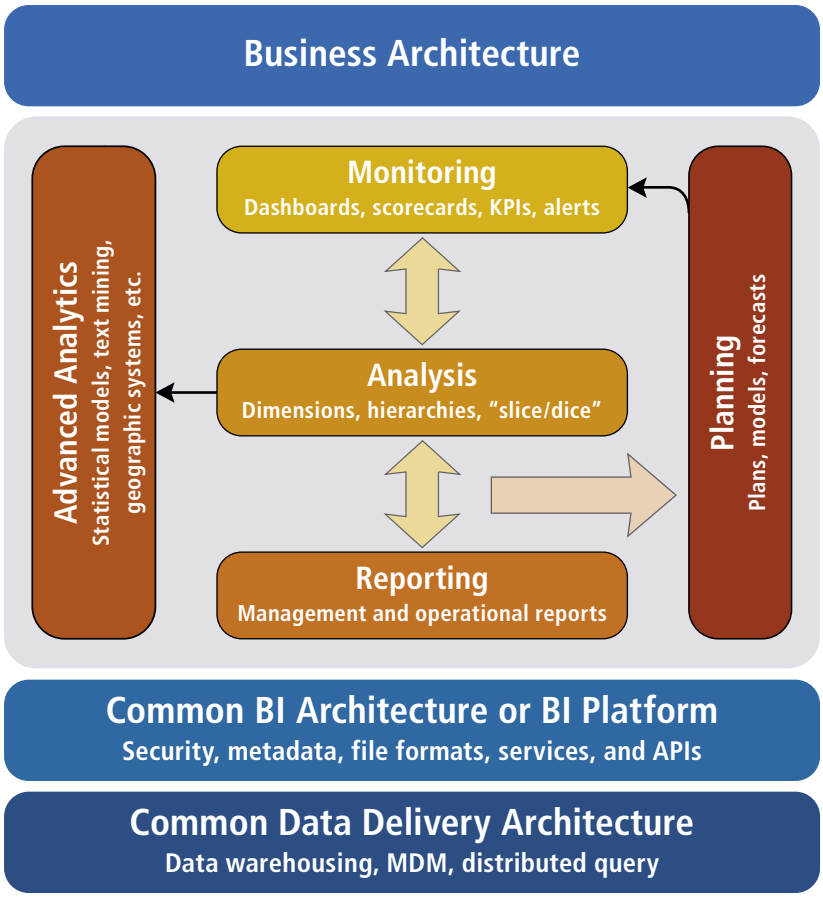


Figure 2. *BI platforms transform once-separate BI tools into an integrated set of applications that are aligned with strategic objectives and run on a single BI architecture built on top of a robust data infrastructure.*

that overhauling the analysis dimension wasn't necessarily going to improve their ability to penetrate the larger market of casual users who found OLAP tools too demanding. Many BI vendors discovered that their strategy of touting the "self-service" capabilities of their analytical and reporting tools was backfiring. Most users, they realized, don't want to create ad hoc queries, build reports, and slice and dice data; they simply want the right data delivered to them at the right time, and only when there's something critical they need to examine. In other words, instead of self-service BI, users want customized delivery of information and insights.

Dashboards and scorecards. So what does customized delivery look like? BI vendors discovered that most users want to monitor key metrics using a simple graphical display that also alerts them when performance falls below predefined thresholds. So vendors began delivering dashboard and scorecard interfaces on top of their analytical, reporting, and planning tools. Dashboards and scorecards let users compare performance to plan at a glance and send alerts about any deviations through user-defined channels of communication (e-mail, pager, cell phone, etc.). When an exception condition occurs, users then drill into the analytical tools to perform a root cause analysis (or pass the alert to their business analyst to do this) and then to reporting tools to examine detailed data and better determine what actions to take. Finally, they use planning tools to create new forecasts, readjust goals, and establish new monthly targets and thresholds.

BI platforms. Today, leading BI vendors are working feverishly to seamlessly integrate the monitoring, analysis, reporting, and planning dimensions of BI. They are redesigning their products to run on a service-oriented architecture so that all tools share common services such as security, metadata, data access, rendering, etc. (See Figure 2.) These BI platforms transform distinct BI products into a suite of highly integrated applications running on a shared set of services within a single architecture. This will enable users to access information in any dimension of BI and then seamlessly move between applications (or layers) as their needs dictate.

5. Advanced Analytics Dimension

While BI vendors rush to offer a BI platform that provides integrated reporting, analysis, planning, and monitoring, user requirements keep moving ahead. There is now a groundswell of interest in exploiting new dimensions of BI, namely data mining, text mining, and advanced visualization. I group these under "advanced analytics." Many companies are now asking, "How can we deliver further value from our DW/BI investments?" and, "How can we query, analyze, and report on both structured and unstructured data?" Advanced analytics are the answer.

Data mining. For years, a small number of companies (mainly in the financial services industry) developed statistical- and rules-based models to detect hidden patterns in large volumes of customer transactions. These "data mining applications"—fraud detection, cross-selling recommendation engines, dynamic pricing, Web traffic analysis and site design, promotions marketing—yield tremendous payback on initial investments and often spell the difference between profitability and disaster for many companies.

Today, vendors are embedding data mining algorithms and rules-based engines in a host of packaged analytic applications to disseminate their benefits to a wider audience of business users without requiring them to be statistical rocket scientists. Vendors are also bolting advanced visualization technologies into data mining–based applications and OLAP tools to accelerate the analysis and interpretation of results.

Unstructured data analysis. BI vendors are also looking to enter the vast untapped market for analyzing unstructured data, which includes text, documents, images, geospatial data, video, Web pages, and so on. Unstructured data constitutes 80 percent of all information within organizations. Vendors who figure out how to marry the distinct worlds of structured and unstructured data access and analysis will garner a lot of attention and business.

Today, there are two basic approaches for combining unstructured and structured data for analysis. One uses search technologies to index structured data, while the other uses data warehousing techniques to extract and load unstructured data into relational databases. The first technique indexes both structured and unstructured data so that both can be queried using keyword searches. The second loads both types of data into data warehouses, where they can be accessed using SQL-based query tools. Both approaches have advantages and disadvantages, and neither is ideal at this point.

Conclusion

It should be obvious by now that BI is a rich and ever-expanding field. I see no limits to the growth of this industry, because it serves the basic need of individuals and organizations to transform data and information into insight and knowledge to drive profitable business activity. Such tasks are quickly becoming the hallmarks of highly agile and competitive companies. Consequently, there will always be a market for new and innovative ways of delivering insight and information to decision makers. So, for those of you in the BI industry, buckle your seatbelts—the ride has only just begun! ■

Wayne W. Eckerson is TDWI's director of research and services. Eckerson has 17 years of experience in the IT industry, and he has covered data warehousing and business intelligence issues since 1994. He is the author of Performance Dashboards: Measuring, Monitoring, and Managing Your Business *(John Wiley & Sons), published in October 2005. You can reach him at weckerson@tdwi.org.*

Verispan Analyzes Massive Pharmaceutical Database with Enterprise BI

Commentary by Michael Haggerty, Director of Business Intelligence Solutions, Verispan

Verispan, a healthcare informatics joint venture of Quintiles Transnational Corp. and McKesson Corp., provides a broad array of information products and services to the healthcare industry, including sales targeting and compensation products; market research audits; healthcare profiles; comprehensive managed care offerings; primary market research; opinion leader mapping; data integration, warehousing, and mining; data analysis and consulting; direct mail; list services; disease management studies; and cost/benefit analyses, among many others. Verispan is the nation's leading provider of patient-centric longitudinal data, with dozens of products used by clients spanning the industry. Headquartered in Yardley, PA, Verispan employs more than 500 dedicated healthcare information professionals. The company's Web site is www.verispan.com.

Verispan Adopts MicroStrategy for Enterprise BI

Verispan chose MicroStrategy for its ease of customization, scalability, reliability, and reusability. Key capabilities include MicroStrategy's ability to work against very large databases, a rich reporting feature set, tight integration with Oracle, the platform's acceptance across the pharmaceutical industry, and flexible deployment (Web or desktop; intranet or Internet).

MicroStrategy supports 20 BI applications that are used both directly by Verispan's client users and internally in support of client deliverables. Vector One®: Total Patient Tracker (TPT) is the first and only service that projects the distinct number of patients taking a therapy or multiple therapies over time, allowing clients to develop optimal marketing strategies based on patient characteristics within specific drug markets or therapeutic areas. This dynamic system allows users to modify market definitions and/or time periods for study. TPT has been in production for three years and is utilized by more than 250 external clients and Verispan employees.

Vector One™: Prescriber Dashboard provides clients with a quick view of key prescribing events such as new patient starts and drug switching/adding. Prescriber Dashboard provides many key prescriber and sub-national reports that help sales operations groups target the right physicians, at the right time, with the right message, which is critical to the success of their brands. Data is available weekly, monthly, or in greater time aggregations. Markets are predefined for easy data evaluation, and every drug available through retail pharmacies is available for analysis. The system is used internally by Verispan to deliver approximately 150 detailed prescriber reports to clients each month.

Pharmacy Cost Analyzer (PCA) is a Web-based pricing analysis tool that compares a client's cash and third-party transactions against 20 other chains and independent pharmacies, which is critical to maximizing revenue. PCA enables users to understand the key pricing metrics within their markets and maximize their competitive advantage. This user-friendly, point-and-click application is updated weekly and maintains a four-week rolling history, giving clients the timeliest access to benchmark data available.

User, Data, and Report Statistics

Verispan's database has four years of data history and continues to grow. There are twice-weekly and once-monthly data loads, and it is available 24/7 to its 1,675 total users. Verispan's team includes 10 MicroStrategy developers.

VERISPAN
THE TRUE MEASURE OF HEALTHCARE

Verispan's database is a 120-terabyte Oracle 9i® data warehouse. MicroStrategy projects currently run against approximately 10 of the 120 terabytes. The warehouse hardware includes seven IBM pSeries UNIX servers attached to EMC storage. Fourteen application servers host MicroStrategy products, including version 8.0.1 I-Server, Web, and Narrowcast Server, across the production, test, and development environments.

According to Verispan, other BI products have difficulty handling their high data volumes at a granular level. Prior to MicroStrategy, only a limited set of skilled developers and analysts at Verispan had direct access to the data, and most reporting was done with custom PL/SQL scripts. This approach was not scalable because most of the code written was not reusable. With MicroStrategy, Verispan has been able to open up the patient-centric data warehouse to a wider group that includes non-technical employees.

Business Wins

Since adopting MicroStrategy, there has been widespread user acceptance among Verispan employees, and several clients have contracted with Verispan to build and host custom data marts with a MicroStrategy front end. MicroStrategy-based product offerings are also a major part of a multiyear contract Verispan was recently awarded with the U.S. Food and Drug Administration, providing the agency with data to monitor trends in the pharmaceutical industry. ■

High-Volume, Real-Time Data Transformations Become Reality at Overstock.com

Commentary by Jack Garzella, Vice President of IT Operations, Data Warehousing, Reporting, & Analytics, Overstock.com, Inc.

Company Overview

Overstock.com, Inc., is an online "close-out" retailer offering discount, brand-name merchandise for sale over the Internet. The company offers its customers an opportunity to shop for bargains conveniently, while offering its suppliers an alternative inventory liquidation distribution channel.

Managing the Data Deluge

With several terabytes of data stored in their enterprise data warehouse, and millions of transactions taking place each day, Overstock was constantly battling to keep the data in its data warehouse accurate and up to date.

To better track and improve e-mail campaigns, and to better understand and manage customers and suppliers, Overstock wanted to establish an IT environment that would allow management to understand at any point in time how well it was meeting its performance target metrics.

> "Having access to key business metrics in real time is no longer a fantasy."
>
> —Jack Garzella, Overstock.com

To gain access to relevant, timely information—information that could be used to help guide business decisions—Overstock realized that it needed to integrate data from disparate data systems and applications. Thus, they set out to find a data integration product that could handle the company's high-volume, highly complex loading and transformation requirements in near real time.

Meeting the High-Performance, High-Volume Data Integration Challenge

After looking at various products, Overstock selected Sunopsis' Data Conductor. Several factors influenced their choice:

First, Overstock really liked the performance gains provided by Data Conductor's clean architecture and approach to moving data. "It is not a typical ETL [extract, transform, load] tool," said Jack Garzella, who runs Overstock's IT operations, data warehousing, reporting, and analytics. "With a traditional ETL tool, we would have been forced to rely upon a proprietary engine to perform our transformations. We didn't like the idea of having to move our data first out of the database, then into the proprietary engine for processing, and then back into the database. This approach—because of the multiple data movements and the large volumes of data we deal with—would have made it nearly impossible for us to gain access to data in near real time."

Overstock realized that by choosing a traditional ETL product, they would not only experience performance issues due to the way the proprietary engine handled data, but they would also be forced to incur additional hardware—adding further costs and complexity to their IT environment. "With Sunopsis' unique E-LT [extract–load, transform] approach, we are able to use the power of our existing RDBMS engine to perform the transformations," Garzella said. Data Conductor's use of business rules to handle integration processes, and the product's ability to handle both batch and real-time integration, were other key factors influencing Overstock's decision to choose Sunopsis.

Simplicity, Performance, and Scalability

Historically, Overstock handled transformation requirements by writing custom code anytime data needed to be modified or moved. This was a time-consuming and tedious task, since the company processes millions of records per day. With data stored and written in multiple formats and files, and with each transaction on the Overstock.com site requiring data to be populated in six to eight different records, data movement was a constant challenge.

"Upon looking at Sunopsis, we realized they were a perfect solution for meeting our high-volume, high-performance requirements, complex transformations, and near-real-time needs," said Garzella. Using Sunopsis' highly powerful ETL product allows Overstock to automatically and efficiently load and transform large volumes of data from source systems directly into their enterprise data warehouse.

Sunopsis—Better Data, Better Intelligence, Better Decisions

Sunopsis Data Conductor's unique architecture makes it possible for Overstock to automatically transform over five million daily transactions into useful data. Using the product, Overstock is able to perform even the most complex data transformations on very large volumes of data without experiencing any performance problems.

"Having access to key business metrics in real time is no longer a fantasy," said Garzella. "With Sunopsis, over 300 Overstock users in the sales, finance, marketing, and merchandising teams are now able to view their relevant data in near real time hourly, daily, or weekly, depending upon their needs—making it easier to make better decisions and better manage the company's bottom line." ■

Legg Mason Wood Walker Gains Competitive Edge in Financial Services Industry

Commentary by Lisa Hayes, Director of Operations, Legg Mason Wood Walker

About Legg Mason Wood Walker

Legg Mason Wood Walker, Inc., is a financial services holding company that provides asset management, securities brokerage, investment banking, and related financial services through its subsidiaries. Asset Management Technology Solutions (AMTS), a wholly owned subsidiary of Legg Mason providing strategic client services, including customized statement packages and fully integrated portfolio systems, is designed to fulfill specific operational and technical demands.

The Legg Mason Challenge

Legg Mason's executive team mandated that its reporting solution provide a competitive edge. There were two specific areas identified where the reporting solution needed to add value:

1. Make the financial advisors more effective
2. Make the information it provides its clients the best in the industry

Legg Mason has more than 1,200 financial advisors serving more than 20,000 clients, all with varying needs. According to Lisa Hayes, director of operations for Legg Mason Wood Walker, "It was a struggle to generate the volume of reports for the large number of accounts that we have. Many programs can't handle the volume."

The Actuate Solution

Once AMTS saw the quality of Actuate's report presentation, the decision to go with Actuate was easy. "Actuate stood out among the other packages we were looking at because of its scalability, flexibility, and the strength of its distribution of the report packages," said Hayes.

AMTS has leveraged Actuate's graphical capability to create reports that are much more visually rich and easier for information consumers to access and digest. AMTS took this a step further, leveraging the Actuate technology and APIs to develop its own portfolio management application that operated in and around Actuate.

The Benefits of Actuate

With Actuate, Legg Mason produces client statements faster and in a more presentable format. Current information is available anytime, rather than a month after the quarter close. The financial advisors can immediately access the reports via the intranet. Cutting down the time it takes to produce the reports and distribute them online to financial advisors has been a significant enhancement to Legg Mason's business.

Today, the financial advisors can sit down with clients and pull up their current holdings, as well as performance gain and loss reports. Brokers use this information to help in the decision-making process. Now the financial advisor can look his client in the eye and say, "This is where we are today."

"We're hearing success stories from financial advisors getting more assets because the clients are happy—not just with the financial advisor, but with the reporting as well," said Hayes.

"It's certainly not there to replace the financial advisor, but we need to have a very competitive tool because the report is something tangible that our clients get every quarter."

Hayes elaborated, "We just recently supplied the reports to someone with non–Legg Mason assets. They saw the reports and had to have them. As a result, the broker brought in another four million dollars in assets."

The Future

As AMTS drives to enhance functionality for its customers, including Legg Mason Wood Walker, it will be proactively enhancing functionality with new Actuate capabilities. Its goal is to continue to expand that functionality and leverage the features in the latest version of Actuate to make Web-based reporting more interactive and richer, both from a client perspective and a financial advisor perspective.

"[We now have] the flexibility and the power to process the volumes that we have without issues, the power to present things graphically and aesthetically without the application blowing up, which we've seen in other programs. It all provides us a competitive edge. We get a lot of positive attention both within and outside of Legg Mason for our quarterly reporting. We feel that we have the premier reporting at this point," concluded Hayes. ■

> "Actuate stood out among the other packages we were looking at because of its scalability, flexibility, and the strength of its distribution of the report packages."
>
> —Lisa Hayes, Legg Mason Wood Walker

Enhanced Visibility and Superior Customer Service at Iron Mountain

Commentary by Paul Tsitsopoulos, Director–Operations Engineering, and Steve Weber, Manager–Operations Engineering, Iron Mountain

Company Overview

Iron Mountain Incorporated is the global leader in information protection and storage services. The company offers comprehensive records management and data protection solutions to more than 90,000 corporate clients throughout the U.S., Canada, Europe, Latin America, and the Pacific Rim.

Business Challenge

As a by-product of numerous years of consistent, profitable growth, Iron Mountain recently discovered several opportunities to improve the efficiency of its "customer on-boarding" processes. The information storage leader needed to bring new customers into its technology infrastructure in a more streamlined manner. Specifically, it wanted to reduce the number of groups involved in on-boarding, optimize the process, and improve coordination among participants.

As part of its new process, Iron Mountain wanted to ensure that tasks were not transitioned from group to group with incomplete or erroneous data, and that assignments that could be handled simultaneously were executed in parallel. Furthermore, the organization wanted to enhance its up-front customer data entry and data management capabilities.

To help achieve these objectives, and create an environment more amenable to superior customer service, Iron Mountain engaged Collaborative Consulting.

Solution

A team of Collaborative business process management and technology experts examined Iron Mountain's existing customer on-boarding processes and identified areas where tasks previously performed in a linear fashion could be done simultaneously. However, to compress the period usually required for on-boarding, Iron Mountain needed access to information in real time. That meant it had to change the way it entered and managed new customer data.

To accommodate these requirements, Collaborative worked with Iron Mountain business and technology professionals to develop a Web application to manage up-front data entry more smoothly. In addition, the team implemented a workflow management system to more efficiently add new customers' digital assets to Iron Mountain's technology infrastructure. The team also established a repository to hold new customer information.

Like many companies that successfully implement BPM projects, Iron Mountain has appreciably improved visibility into executing processes, as well as enhanced reporting of current and historical process information.

Benefits

As a result of its BPM solution, Iron Mountain is able to bring its customers on board much faster than before. In addition, employees who used to manage time-consuming on-boarding tasks are freed up to work on other projects. This ability has enhanced revenue collection and customer satisfaction while reducing operating costs.

And because the team automated a host of processes that had previously been performed manually, notably data validation and batch loading of data, Iron Mountain customers enjoy a more consistent experience. Supporting processes that are still done manually are more streamlined and consistent throughout the organization. Moreover, like many companies that successfully implement BPM projects, Iron Mountain has appreciably improved visibility into executing processes, as well as enhanced reporting of current and historical process information. As a result, Iron Mountain managers can spot bottlenecks as soon as they occur, and sometimes even before they occur, eliminating many problems and improving manageability.

The solution also facilitates continuous improvement. In the future, Iron Mountain will benefit from enhanced development productivity resulting from the BPM architecture Collaborative created. This framework will enable the organization to automate as much as 70 percent of the manual tasks remaining in its business process. Iron Mountain will also be able to use the data it collects over the next several months to establish actual performance metrics for various tasks. As such, it will be able to create service-level agreements, measure them precisely, and act quickly when they are missed. In summary, Iron Mountain is able to provide exceptional service to its customers—particularly new customers—that will only increase as additional BPM tools and techniques are employed. ■

Dean Hoff Consulting Improves the Analytics Capabilities of a Major Direct Marketing Agency

Commentary by Dean Hoff, Dean Hoff Consulting

The Challenge

Located in Minneapolis, MN, Dean Hoff Consulting helps clients develop and implement analytic strategies. This includes building business intelligence applications utilizing software and other tools, then showing clients how the analytics can interact with their current processes. While working with a major direct marketing agency, Dean Hoff ran into a problem when developing predictive models. He explains:

> Part of the process required us to apply model scores to a large demographic file containing over 100 million records. We developed the models and wanted to implement them, but in order to do this, we needed to be able to score the records. We went to the company's internal resources to determine if they had the capability to accomplish this. The size of the file was too large, so they told us to outsource this. After discussing possible solutions with a third party, the costs would have been too prohibitive. I knew there had to be a way that we could do this internally.

The Solution

After conducting extensive research, Hoff decided to test DMExpress, the high-performance data transformation product from Syncsort Incorporated. He found that DMExpress allowed him to calculate a model score as well as score a large number of records in a short amount of time. Hoff commented, "We now take a historical marketing file and append the demographic information using DMExpress. Basically, we're matching the marketing file with a large demographic file using the join functionality. Then we develop the model using the appended file, take the results of that model, and score all of the demographic records we have." Once this is completed, the company has a breakdown of everyone in a geographical region who was contacted and whether they ordered a product. The information can then be used to improve the response rate of a direct marketing campaign by reaching out to consumers who are more likely to make a purchase.

The Benefits

One of the key benefits of DMExpress for Hoff is its speed. "Instead of taking days to run the aggregations, I'm able to complete the process in minutes with DMExpress," he says. It also provided Hoff with the functionality he needed to process data in other areas of the application. "I initially integrated DMExpress into the application to just handle the scoring, but because of its join capability, I was able to use it to do the data append, too. With the aggregation component, I can actually do some profiling with the data. Because of this enhanced functionality, I'm using it beyond the initial intent." ■

Overview

Organizational Profile

- Develops and implements analytic strategies for companies
- Specializes in the direct marketing industry

Business Needs

- Minimize the elapsed time of a business intelligence application
- Perform complex aggregations on a large amount of data

Environment

- Windows XP

Benefits

- DMExpress processed 100 million records in less than two hours
- Using the Advanced Data Management component, DMExpress completed the complex aggregations much faster than anticipated, taking minutes instead of days

> "Instead of taking days to run the aggregations, I'm able to complete the process in minutes with DMExpress."
>
> —Dean Hoff, Dean Hoff Consulting

Global Hyatt Corporation Consolidates 200+ Business Units

Commentary by Gebhard Rainer, Vice President for Hotel Finance and Technology, Hyatt International Corporation

Summary

Hyatt uses Hyperion for global consolidation of 200+ business units, reducing month-end closing time from 15 to 2 days, and cutting the forecasting process from five months to six weeks. Hyperion also assists with improved data accuracy and compliance. Palladium, AnswerThink, and Hyperion assisted with implementations.

About Global Hyatt

There are 211 hotels and resorts (over 90,000 rooms) in 44 countries around the world operating under the Hyatt®, Hyatt Regency®, Grand Hyatt®, and Park Hyatt® brands.

Spreadsheet-Based System Obsolete

With the establishment of Global Hyatt Corporation, there arose a need to more quickly and efficiently consolidate the hotels' financial data and streamline the planning processes throughout the business units worldwide. "Our manual aggregation of financial data had become far too time-consuming," said Gebhard Rainer, vice president for hotel finance and technology at Hyatt International Corporation. "We consolidated operating results from a management perspective, to gauge our hotels' operating performance on a monthly basis, which typically took 15 days. And then we had to consolidate the results from our legal entities into the corporate structure, which took anywhere from 45 to 75 days. Plus, our old budgeting process took anywhere from three to five months.

"In addition to the amount of time needed, we had issues with the accuracy of the data," Rainer continued. "Using the Excel spreadsheets for consolidation was a highly error-prone process. So putting the two concerns together, we needed a system that would automate these processes and allow us to do global consolidation across all of the business units.

"The second equally, if not more important, driver in our purchasing decision was that we looked at many *Fortune* 100 companies, and found that a great majority use Hyperion," Rainer adds. "For these reasons, combined with our shared vision of business performance management, we felt Hyperion offered the right solution for us."

Hyperion Standard Assists with Compliance

Global Hyatt Corporation launched Hyperion company-wide in August of 2004. Says Rainer, "Some of the selection process, in terms of choosing which applications to roll out first, was dictated by our business processes. We knew Global Hyatt Corporation was going to become an active company in January of 2005. That meant that for the first quarter, we would need to have systems in place that would allow us to do our global consolidation accurately and in a timely fashion. With Hyperion, we easily met that objective." Hyatt now pulls and consolidates its source financial data from over 230 locations, where it resides in an Oracle database and several local data marts. "Every Hyatt-managed hotel in the Americas, Europe, and Asia is using Hyperion for budgeting and forecasting, and Hyperion for financial consolidation," Rainer notes. "Since we're a global corporation, we must deal with a lot of variation in our general ledgers and accounts—we have to maintain an extended set of accounts for local requirements. And we actually follow both U.S. and local GAAP [generally accepted accounting principles]."

Faster Closes, Higher-Quality Data

Additional benefits of Hyatt's Hyperion implementation include a greatly reduced time frame for financial consolidation and significantly improved accuracy in the forecasting process. Rainer says this reduction has been dramatic: "Thanks to the Hyperion application suite, we've made a huge improvement in our corporate close and reporting, which now is within 30 days from quarter-end, over the previous 50 to 75 days."

Rainer says that the single most important benefit of adding the Hyperion solution—including its powerful drilldown capabilities—has been that Hyatt executives can now track their data directly to its source.

Financial Scorecards in the Future

"In the years ahead, we will extend our Hyperion solution into the whole business intelligence area, in order to make other, non-financial information available to Global Hyatt Corporation management," Rainer predicts. "That will lead us to a scorecarding approach, where we will be able to compare the results with targets and key performance indicators, and get almost real-time feedback on our financial results from the business units. In this way, Hyperion will help us respond in a much more proactive fashion than has ever before been possible." ■

> "We have achieved our goals in the area of financial consolidation and global financial planning, significantly reducing the time needed for both while greatly improving the accuracy of our data."
>
> —Gebhard Rainer, Hyatt International Corporation

Wimbledon Tennis Championships Go On Demand

Commentary by Rob McCowen, Marketing Director, All England Lawn Tennis and Croquet Club

The Wimbledon Tennis Championships are an established event in the British sporting season: A unique, quintessentially English occasion that conjures up images of world-class tennis played on immaculate grass courts, good-humored spectators and, of course, the interruptions of the British weather.

A World-Class Event from a Small, Private Club

The All England Lawn Tennis and Croquet Club stages the Championships and is very proud of its unique reputation. For two weeks every year, this small club's infrastructure scales up exponentially to accommodate the demands of the world's sporting media, half a million spectators onsite, over half a billion television viewers, and around four million unique users of the official Wimbledon Web site.

Drawing the Eyes of the World to Wimbledon

Real-time information is important in attracting and retaining new players and audiences. Players use it to improve their game, and the information provided on the spot to international television and broadcaster graphics systems gives color and depth to the viewers' experience. Broadcasters can access the Commentator Information System to enrich their commentaries with statistical analysis and comparative performance data.

Tennis fans around the world take a "virtual seat" at www.wimbledon.org and feel part of the action. The IBM On Demand Scoreboard delivers live point-by-point scores directly to the desktop, and the new Stroke Tracker ball-tracking system is available online for the first time this year, enabling users to compare players' games shot by shot.

The Wimbledon Web site is vital to stimulating interest in the Championships, and IBM Business Consulting Services works with the club to ensure that the site generates year-round business. The online shop is particularly popular in the U.S., China, and Japan. Rob McCowen, marketing director at the Championships, says: "Previously, we developed business opportunities based on awareness created through our TV coverage. Increasingly, new business opportunities arise through interest generated by our Web site. This is particularly relevant in the U.S. and Asia. Companies in Japan, for example, are learning more about Wimbledon, leading to new revenue streams."

Improved Decision-Making and Efficiency

A new management solution was used in last year's Championships to support player accreditation, manage players' guests, and coordinate other player facilities such as the provision of limousines. This year, press, radio broadcasters, and photographers will be managed through the solution. Eventually it will cover everyone on site at Wimbledon who is not a paying ticket holder. Administration overhead will be reduced as records are no longer prepared manually, and there will be less duplication of effort. Managers will also gain access to better reporting facilities, improving security, decision making, and efficiency.

Continual Innovation in an English Country Garden

Players, tennis fans, and the Lawn Tennis Association all benefit from innovations that reach out to global audiences through new content channels, optimize revenue streams through e-commerce or awareness raising, and increase the efficiency with which the Championships are organized and run. ■

On Demand Business Benefits

- Drives business innovation through the use of new market channels
- Delivers cost-effective scalability, enabling a small club to host a global sporting event
- Increases resilience
- Meets increasing audience expectations

On Demand Scalability Defined

- A seamless, high-performance solution that flexes to meet business demands
- No large, upfront capital investment
- Wimbledon does not own, manage, or maintain the infrastructure, leaving it free to focus on its key business

"Previously, we developed business opportunities based on awareness created through our TV coverage. Increasingly, new business opportunities arise through interest generated by our Web site … Companies in Japan, for example, are learning more about Wimbledon, leading to new revenue streams."

—Rob McCowen, All England Lawn Tennis and Croquet Club

Data Warehouse Appliances in High-Tech Manufacturing

A Better-Performing Data Warehouse for the Former Cost of Annual Maintenance

By Mark Theissen, VP Professional Services, DATAllegro

In order to remain competitive, high-tech manufacturing companies must manage their costs, especially their IT costs. The following scenario was recently faced by a leading high-tech manufacturer in the computing industry; we'll call the company HTM, Inc.

HTM had great success in their implementation and use of data warehousing to improve production and quality and drive competitive advantage. The success of their data warehousing efforts resulted in increasing numbers of users and increasing demands for data. Being a global company, HTM's data warehouse was accessed worldwide. Batch windows for daily updates were becoming smaller while the workload was increasing. The data itself continued to grow as production grew at double-digit rates. In addition, users were demanding that more and more data be retained for historical analysis. The data warehouse was reaching its capacity.

The first option HTM explored was expanding their existing infrastructure to meet increasing demands. But the existing infrastructure was expensive, and it became apparent that the price tag associated with increasing the size of the data warehouse environment would be cost-prohibitive. HTM needed an alternative that would add more capacity, increase scalability, and *reduce* costs instead of increase them.

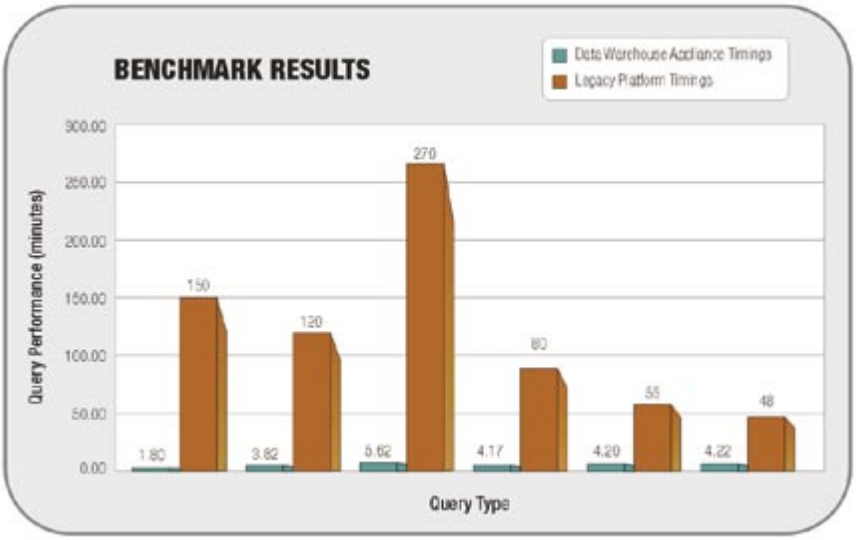
As a leader in high technology, HTM began to look at emerging technologies to meet its data warehouse infrastructure needs. Data warehouse appliances quickly became a point of focus, because they offer a number of benefits:

- A purchase price close to what HTM had been paying annually for the maintenance of its existing data warehouse infrastructure.
- Increased query performance over the existing environment.
- Load times that were hundreds of times faster than current methods.
- Room for growth and the ability to expand the appliance without replacing it.
- A lower total cost of ownership.

All of these benefits translated into an attractive solution that needed to be proved out before it could be purchased. A proof-of-concept project was run that exercised the appliance in terms of throughput, concurrency, and scalability against a mixed workload of queries. The results were outstanding, with load times over 200 times faster than the existing environment. Queries ran from 10 to over 100 times faster than the existing data warehouse.

Best yet was the fact that implementing and maintaining the appliance was a straightforward process. The infrastructure (i.e., database, hardware, OS, software, and storage) was pretuned and configured for data warehousing. The existing data warehouse data model could be used with no modifications. DBA and system administration time was reduced as the number of indexes was minimized, space management was automated, aggregations were not required, and query tuning was greatly reduced.

Today, many companies have the same challenges that HTM faced. Usage of the data warehouse is increasing, the data warehouse is becoming mission critical, and the variety, velocity, and volume of data is increasing. Data warehouse appliances can drive IT costs down while delivering new levels of performance and scalability at a price that cannot be matched by traditional data warehouse infrastructures. Indeed, data warehouse appliances can enable analytics that were previously impossible or unaffordable. ■



It should be noted that the query results from the existing system were run while the system was running concurrent processes, and the queries run in the data warehouse appliance environment were run individually. That said, the queries on the data warehouse appliance were run without any tuning or indexing, while the queries on the existing system had been tuned and optimized by a team of people over several months.

A business intelligence or data warehouse implementation can be a formidable undertaking. In these pages, leading business intelligence and data warehousing solution providers share their answers to the questions they hear often from industry professionals. Tim Feetham, an independent consultant, provides his analyst viewpoint to each Q&A.

Actuate Corporation

Introducing performance management reporting to a diversely skilled workforce is incredibly expensive and time-consuming. What is the best strategy for driving adoption of new reporting applications across the enterprise?

Large user communities, as a whole, do not want to learn new technologies. They prefer to use the productivity applications to which they are accustomed. Therefore, in order to ensure that operation performance management solutions are widely used, we recommend deploying them in familiar means such as personalized, interactive Web applications and fully functional spreadsheets. This way, we immediately overcome the first objection to new technology. Instead, users say: "I knew how to use it when I opened it."

Analyst Viewpoint

In performance management, getting relevant feedback to the folks who can make a difference is essential. Organizations that undertake this task must also understand that what is relevant today will change tomorrow. These organizations face two issues: how to generate broad adoption and how to stay flexible. These needs point to Web-based reporting technologies that deliver reporting, visualization, and analysis tools, plus seamless spreadsheet integration under a unified but customizable interface. Users will be quick to adopt this technology, and support organizations will be able to focus more on tailoring the product to the pressing business issues at hand.

Request more information about Actuate Corporation

Business Objects

Why is EIM so important to BI?

It's not enough for business intelligence (BI) software to simply produce a pretty chart or report. It has to deliver information to people that is credible and accurate—information that people can trust. Look for BI vendors with integrated products, services, and partnerships that can help you deliver enterprise information management (EIM). EIM is the strategy, practices, and technologies needed to deliver a comprehensive approach to managing disparate data in order to drive performance. EIM requires robust data integration and data quality capabilities that are tightly linked to the BI platform.

Analyst Viewpoint

Often organizations spend considerable money on business intelligence (BI) technologies, but don't support their purchases with sufficient planning. Such action will likely result in poor data quality, a lack of timely access to requisite information, and a sense that BI was a poor investment. Savvy organizations will select their BI technologies within the framework of an enterprise information management (EIM) strategy. EIM acts as an umbrella program for data architecture, data governance, data warehousing, and information delivery initiatives, as well as an organizing principle for information workers. When BI technology is deployed in such an environment, the payoff is dramatic.

Request more information about Business Objects

Collaborative Consulting

User adoption of our business intelligence tools is low. What's wrong?

There are two possible reasons for disappointing BI tool adoption: Either the tool is a poor fit, or users don't need or trust the data it uses.

A poor fit may result from one or more of the following reasons:

- IT chose the tool without user involvement or adequately defined requirements.
- A tool evaluation was performed without guidance from experienced practitioners.
- Someone has used the tool previously and figures it will work well again.

Of course, if users don't trust or need the data provided, even the best BI tool won't help.

Analyst Viewpoint

Business intelligence (BI) tools by themselves do not deliver business value to the organization, and without business value, users have little incentive to learn these tools. Having high-quality data that holds the potential answers to key business questions and that is stored within an easily understood structure is a prerequisite for broad BI tool adoption. Further, BI tools must fit the needs of different types of users in order to gain wide adoption. Wide adoption requires the involvement of both casual users and analysts. Organizations without deep experience in these areas will be well served by working with knowledgeable consultants.

Request more information about Collaborative Consulting

DataFlux Corporation

What is the role of data quality in a customer data integration (CDI) initiative?

Put simply, effective CDI requires powerful data quality technology. When creating a "single view of the customer," organizations need a technology that can standardize, verify, and correct data across sources. In addition, data quality technologies provide matching technology—also called identity management. This helps resolve instances of the same customer record across sources—and allows companies to understand the total value of every customer. CDI vendors have typically partnered for data quality technology, but by using a data quality solution as the foundation of CDI, you can realize a faster time-to-value from CDI initiatives.

Analyst Viewpoint

There can be no more essential data quality effort than in the area of customer data integration (CDI). CDI is tricky. It requires that different departments, such as sales, service, and accounts receivable, share key pieces of customer data. Given that this data reflects the quality of the interaction between an organization and its chief revenue source, poor customer data quality will lead to a deleterious effect on the bottom line. Although individual departments may have specialized data needs that arise from unique customer interactions, well-managed deployment of CDI technology can ensure that shared data is consistent, accurate, and timely.

Request more information about DataFlux Corporation

DATAllegro, Inc.

What is the difference between Ingres and Postgres?

Ingres emerged circa 1974 under Michael Stonebraker and Eugene Wong at UC Berkeley. In 1980, Ingres and Oracle entered the commercial world as the two leading relational database management system (RDBMS) products. Ingres is the progenitor of other RDBMS products such as Informix, SQL Server, and Tandem's NonStop, among others. Stonebraker later developed Postgres, a derivative of Ingres, as an object-relational DBMS (for unstructured data types). Postgres remained open source, but Ingres had engineering oversight since its commercial availability. Computer Associates bought Ingres in 1990, and it is the database within CA products. Ingres is ISO 9001 certified.

Analyst Viewpoint

Michael Stonebraker released Postgres, an object-relational DBMS, to the open source community in 1986. It has since gained wide acceptance in that group. Ingres, which he helped create in 1974, made its reputation as a leading commercial RDBMS. Computer Associates purchased Ingres in 1990. CA made it its key database product and encouraged its customers to move to Ingres for Y2K. Later, CA released Ingres into the open source community. Although both products are now in the open source community, Ingres has benefited from years of market discipline, making it an ideal engine for low-cost database appliance technology.

Request more information about DATAllegro, Inc.

Hyperion Solutions Corporation

Why is having a unified business performance management system critical to the success of today's enterprise?

This is being driven by the need to progress from operating tactically to taking a more strategic perspective. It is no longer sufficient to simply report results that are not tied to strategic objectives or business plans. Enterprises need to link their strategic objectives with operational goals. To implement, businesses must report and analyze financial and non-financial information in a coordinated fashion. This then allows them to monitor and compare performance to plans, adjust plans to respond to changes in the business landscape, and perform advanced planning by modeling potential scenarios. A unified BPM system enables this movement to strategic analysis by integrating a BI platform with a suite of financial applications to deliver tightly integrated capabilities to the business user.

Analyst Viewpoint

A business performance management (BPM) initiative that provides a single set of key metrics (leading indicators) to upper management does not have anywhere near the return of a unified BPM initiative. Only naive managers ignore the fact that individuals at all levels will respond to changes in how they are measured. A unified BPM initiative that includes the right BI technology will measure employees and business partners with five to seven key metrics that support the goals of the enterprise. These individuals must also be the ones to affect these measures. Unified BPM must accommodate new metrics as the environment changes.

Request more information about Hyperion Solutions Corporation

Please turn to page 36 for more Q&A articles

Simon & Schuster Accelerates System Performance and Empowers Reporting Users

Commentary by Paul Zanis, Director of Corporate Data Architecture, Simon & Schuster

When Simon & Schuster brought its reporting problems to HyperRoll, the company wasn't expecting the problems to be solved overnight… but they never imagined they would be resolved within in a matter of days.

"We were very surprised with how fast the HyperRoll solution was up and running," said Paul Zanis, director of corporate data architecture at Simon & Schuster. "The entire system was ready to go after a couple of weeks, and the Business Objects implementation took less than a day."

In a short time, HyperRoll had helped Simon & Schuster resolve two long-standing business problems, reducing its query response times and enabling the company to integrate data from both historical and operational sources.

Solving a Critical Business Problem

Simon & Schuster, a global leader in the field of general interest publishing, is using Business Objects, as well as the company's own Microsoft .NET portal, to access and analyze data. With the need for information at an all-time high from its nearly 1,000 reporting users, Simon & Schuster found that burgeoning data volumes and the demand for increasingly detailed reports had slowed system performance dramatically.

For example, sales representatives running basic sales and inventory reports frequently had to wait an hour or more for a report to run. Reps found that they were planning their day around the reports—sometimes even asking internal data analysts to run reports and e-mail them the results. Simon & Schuster needed a way to accelerate system performance and empower its reporting users to get the information they needed without IT support.

Simon & Schuster was also anxious to better leverage the wide range of information stored not only in their data warehouse, but also in their many transactional systems. The company wanted to be able to integrate operational data with historical information from the data warehouse to help the company make better-informed operational decisions. For example, if the company had a book on back order, they needed a system that would alert the right decision makers to the problem and help them quickly prioritize customer allotments based on customer value and profitability metrics.

HyperRoll's ability to access data straight from source systems has also enabled Simon & Schuster to pull data from other systems that would have been hard to access on a regular basis by business users, enabling knowledge workers to use a broader set of metrics and enjoy deeper analytic functionality.

"Using HyperRoll, we can do things like analyze how quickly we're fulfilling orders, and whether that time period is too long, or even too fast, depending on the customer requirement," Zanis said. "This kind of analysis depends on being able to see transaction-level data from the warehouse, which is greatly facilitated by HyperRoll."

> "The most important benefit we've seen from using HyperRoll is that our sales people can now get the data they need, right when they need it. Faster access to information has truly had a significant impact on our business."
>
> —Paul Zanis, Simon & Schuster

The HyperRoll Solution

Simon & Schuster selected HyperRoll to reduce its query response times and allow the company to use data from both historical and operational sources in its analyses. By using HyperRoll in conjunction with its existing reporting solutions, Simon & Schuster now has lightning-fast access to large data volumes and is able to perform operational analysis as critical business events are happening. For example, the sales report that had taken more than an hour to complete is now returned in just five seconds.

The HyperRoll solution is also transparent to end users. For Business Objects users, the HyperRoll cube looks merely like another available "universe" of data. For .NET users, HyperRoll appears as a flat table. In both cases, end users needed no training to begin taking advantage of the technology as soon as it was deployed. They continued to use their existing reporting solutions… but with greatly accelerated results. ■

Egg Bank Improves Customer Data

Commentary by Egg Bank PLC

Challenges

The world's largest purely online bank, Egg PLC offers a range of financial products and services (credit cards, spend cards, insurance, mortgages, and savings plans) to more than three million customers in the UK. Because Egg is an Internet-based operation that doesn't have a branch network with physical buildings, almost all interaction with customers is done online, and much of it is automated. It's therefore especially important for Egg to have accurate customer information.

Previously, different groups within Egg relied on a number of sources for customer data, resulting in information that was inconsistent, inaccurate, and took up to six weeks to obtain. Egg decided that it had to do something to address this problem.

Solution

In order to ensure that Egg had an accurate, single source of customer information, the bank decided to establish a customer data warehouse (CDW) that would serve as a consolidated, centralized information source.

Egg designed its CDW carefully, ensuring that its infrastructure would meet the service-level requirements of its internal users. Their scalability, reliability, and performance requirements were all factored into Egg's evaluation and planning process. For the CDW database itself, Egg selected Oracle. Oracle features such as parallel query, materialized views, and partitioning would provide Egg with the speed of information that users were seeking.

Data that will eventually populate the CDW is first retrieved in its raw form from a variety of internal and external sources. It's then cleansed, matched, and imported into the CDW. Next, specialized data marts are populated from the CDW for internal departments. Users then analyze the information using a variety of end-user tools, including Oracle Discoverer.

The information held in the CDW is refreshed in near real time. "If someone visits Egg.com, applies for credit with us, is accepted for credit and issued a card, by the time the card is issued, we would see that person in our data warehouse," says Egg head of data Jay Parmar.

> "Previously, users were satisfied with a six-week data latency … We're now doing marketing campaigns using data that is one week behind, but have the capability to conduct daily campaigning with data only 24 hours old."
>
> —Jay Parmar, Egg Bank PLC

Egg's CDW is currently about 2 terabytes in size and is growing by approximately 10 gigabytes each month. Egg campaign managers, creative teams, and phone sales teams all use it to access accurate and current customer information. "If we didn't have this CDW," notes Parmar, "we couldn't do the 120 campaigns per month across 6 channels that we currently conduct."

Results

The CDW has significantly improved the accuracy of the customer data Egg uses. Parmar says, "Previously, users were satisfied with a six-week data latency, and were doing marketing campaigns using data that was six weeks behind. We're now doing marketing campaigns using data that is one week behind, but have the capability to conduct daily campaigning with data only 24 hours old." But the appetite of Egg's internal customers has been whetted, and they are asking for data in 12, 6, or even 4 hours. Egg is consequently headed towards an operational data store—a real-time data warehouse.

Linking customer information to financial and transactional data has proactively eliminated potential problems. As Parmar explains, "The CDW also provides a daily credit decisioning capability to manage risk. For instance, if someone has an Egg credit card, we are able to make a decision using the data in the data warehouse on a transactional basis to determine whether we need to speak to that person or if they are just slightly overdue."

Egg's customers ultimately benefit most from the CDW. "If we didn't have a data warehouse, we would be severely restricted in the service we provide to customers," says Parmar. "At Egg, we put customers first and help them to make the most out of their money. The data warehouse plays quite a strong part in providing the information and data necessary to help our customers make choices about their finances. Outwardly, the customers don't know that we use our CDW to help them, but there's no question of its positive impact on our external customers." ■

For Further Information

http://www.oracle.com/solutions/business_intelligence/dw_home.html

ALSTOM Power Fuels Success with Customer Data Management

Commentary by Daniel Teachey, Corporate Communications Director, DataFlux Corporation

The Business

ALSTOM is a global leader in energy and transport infrastructure. The company serves the energy market through its activities in power generation, and the transport market through its rail and marine activities. ALSTOM employs 75,000 people in over 70 countries worldwide.

The Challenge

For any company with operations across multiple locations, it is difficult to get a true view of the enterprise across geographic and corporate boundaries. ALSTOM Power, a division of ALSTOM, has multiple business systems—including enterprise resource planning (ERP) and customer relation management (CRM)—that contain customer information. Previously, this information was not standardized, and contained duplicate data within and across these systems.

ALSTOM's global corporate strategy of "One Face to the Customer" drove ALSTOM Power's data warehouse group to develop a "customer information broker." This broker provides a single representation of each customer identity, along with a cross-reference to each business system's source record.

For ALSTOM, the customer information broker does not create new information. Rather, it systematically generates a single customer identity based on a modular-designed architecture. This modular design includes data-driven relationships, application-oriented business rules, and user-selected overrides, all of which define how each data element in the customer identity record is populated. For instance, the customer name comes from system A, the street address comes from system B, and the postal code comes from system C. To accomplish this, ALSTOM needed a solution that could:

- Standardize each customer record while maintaining the original value
- Group similar customer entries to remove duplication
- Match according to multiple criteria to increase the potential of positive matches
- Define parent-child relationships, enabling reporting at any customer hierarchy level (holding company, operating company, plant/site, etc.)

"Our data quality issues aren't any worse than other organizations our size, and we have just as much difficulty getting a clear picture of our information," said Mike Sykes, U.S. data warehouse manager at ALSTOM. "To achieve 'One Face to the Customer,' however, we needed to increase our level of data quality dramatically."

The DataFlux Solution

ALSTOM Power selected dfPower Studio, a data management solution from DataFlux, to help build a more comprehensive view of its customers. dfPower Studio offered ALSTOM Power a variety of features, including comprehensive data standardization and de-duplication capabilities, USPS address verification, prebuilt rules that streamlined initial deployment, and transportable rules logic that enables the transfer of rules to other dfPower Studio installations. The architecture of the DataFlux product suite also provides the scalability that will be needed as ALSTOM's data-quality activities mature.

"The first iteration of our customer information broker utilized 25,000 customer records from four business systems," Sykes says. "With dfPower Studio, we created a predictable and repeatable process for generating multiple matching scenarios across our business systems. This was a key feature, as it allowed us to select the highest positive match, which in turn reduced our de-duplication effort."

The Results

"dfPower Studio has performed beyond our expectations and is an integral component of our customer information broker infrastructure," Sykes says. "The software has an intuitive GUI and allows us to customize the underlying standardization and matching logic to meet our needs."

With dfPower Studio, ALSTOM Power has more accurate customer data, resulting in more accurate reports of customer activity. For example, customer revenues are now reported in the 97 percent accuracy range—an achievement that would be almost impossible without a data management solution.

"dfPower Studio allows ALSTOM to view customer information with greater accuracy than ever before," Sykes says. "This translates into real business benefits for ALSTOM, our customers, and our vendors." ■

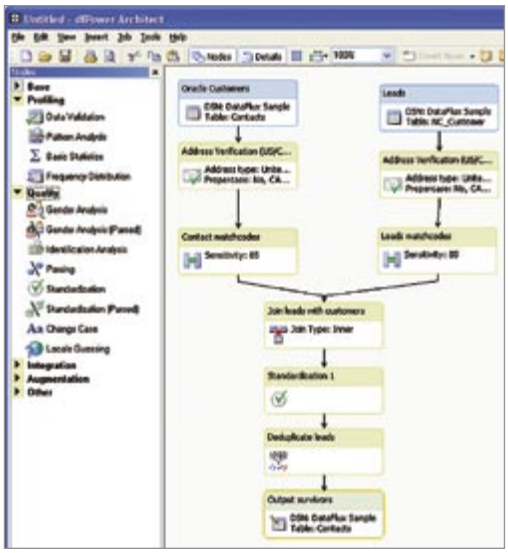


Figure 1. *An easy-to-use workflow builder allows users to perform complex routines, such as address verification, data standardization, and de-duplication, in a single pass of the data.*

Seagate Technology Improves Data Consistency and Eliminates Manual Coding

Commentary by Carin Komer, HR Data Hub Program Manager, Seagate Technology

The Challenge

Seagate Technology is the world's leading manufacturer of hard disk drives for enterprise, consumer, desktop, and mobile systems. With more than 44,000 employees in seven countries across Asia, North America, and Europe, an effective human resources data infrastructure is vital to Seagate's ability to manage personnel acquisition, compensation, and benefits across the globe. Seagate recognized the need for a next-generation HR data infrastructure—an HR data hub—that would reduce integration costs, improve data consistency across multiple units, enhance data availability, and provide cost-effective flexibility.

Over more than a decade, Seagate had knitted together an HR patchwork of legacy and third-party outsourced applications for such HR business functions as payroll, benefits, headcount, stock options, and more. As Seagate's HR system of record, an Oracle HRMS application feeds five localized legacy data stores (homegrown and packaged software atop Oracle databases), which in turn feed more than 100 downstream applications around the world.

Seagate struggled to maintain more than 100 custom-coded integration touchpoints between multiple data stores and a blend of homegrown and packaged downstream applications. The company bore a high cost burden to maintain this intricate system with custom SQL coding, and time-consuming development chores often resulted in a backlog of work for Seagate's human resources information technology (HRIT) developers and frustration for business users.

As applications were added, updated, or removed, inherent problems with inconsistent data and redundant extraction logic multiplied. In addition, the system took up to 48 hours to reflect new information to end users—an unacceptable time lag in a fast-moving global marketplace.

To reduce costs and streamline HR data distribution, Seagate envisioned an HR data hub that would:

- Slash time and costs of custom-coded SQL integration
- Improve data consistency across multiple systems
- Reduce Oracle HRMS integration touch points from many to one
- Accelerate data distribution and reduce latency
- Enable legacy systems retirement to reduce operational costs

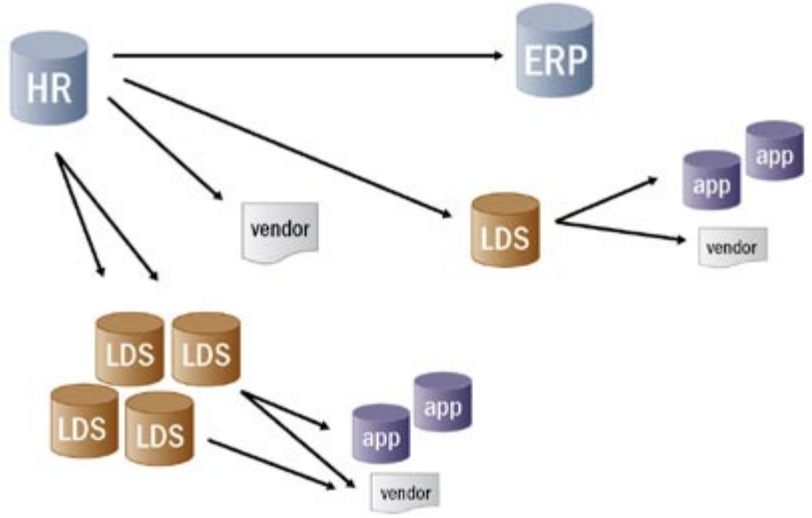


Figure 1. *The complex legacy point-to-point HR data infrastructure was difficult and expensive to manage.*

The Solution

Seagate recognized an opportunity to capitalize on its success using Informatica® PowerCenter® for data warehousing by extending PowerCenter into the operational arena to execute complex data integration among the myriad enterprise applications requiring HR data. Seagate envisioned a concept similar to the customer data hub, designed to create a single view of a customer by integrating information from disparate sources.

Seagate chose PowerCenter for its proven ease of use, mapping templates, and reusability of definitions across projects and platforms. A fast learning curve was particularly important. Although Seagate had highly skilled SQL developers and Oracle database administrators, none of its 10 team members had experience with PowerCenter.

Moreover, Seagate had set an aggressive six-month timeline to go live with the first phase of its data hub and had no budget for external consultants. With a globally distributed development team, Seagate could not afford months of training for its programmers to master a complex tool.

In February 2005, Seagate embarked on a four-phase project that would replace the complex point-to-point data integration network with a single data hub based on the Oracle 9i database. PowerCenter would be deployed to serve two integration functions: 1) extract data from the Oracle HRMS system of record application into the Oracle 9i data hub, and 2) distribute data from the hub to more than 100 downstream applications.

Secure Data Infrastructure

Seagate took advantage of code and version control, migration management, group/folder permission standards, and other PowerCenter security features in building a three-tier integration environment consisting of development, testing, and production. With PowerCenter's component-based reusability, 4.8 full-time equivalents (FTEs) collaborated on parallel activities across the three tiers, including defining data models, installing the development and test environments, and developing hub and interface mappings.

The compartmentalized approach enables Seagate to govern who has access to what code and data to protect employee privacy, abide by best practices in the COBIT IT governance framework, and provide auditability to meet Sarbanes-Oxley and other regulatory requirements.

Enhanced Performance

The team customized a PowerCenter template to build in granular field-level change detection, by which PowerCenter would extract and move only data that had been changed since the prior mapping run. In the development environment, the reduced data volumes enabled by field-level change detection slashed load time from 2.5 hours to 15 minutes compared to previous time/date detection techniques, and minimized performance impact on source, targets, and the Seagate network.

Rapid Deployment

In less than six months—on time and within budget—Seagate went live with the first phase of its HR data hub deployment, covering 10 downstream applications. As it progresses through three additional phases in 2006, Seagate will have consolidated into its HR data hub roughly 100 integration touch points, and virtually eliminated the expensive SQL coding needed to adapt and maintain the proliferation of integration points.

"With PowerCenter, we're able to replace a high-maintenance legacy infrastructure with a high-performance data hub that improves HR data consistency across the globe," said Carin Komer, HR data hub program manager. "We'll save about $1.2 million over three years just by eliminating all the manual SQL maintenance and development chores of the old system."

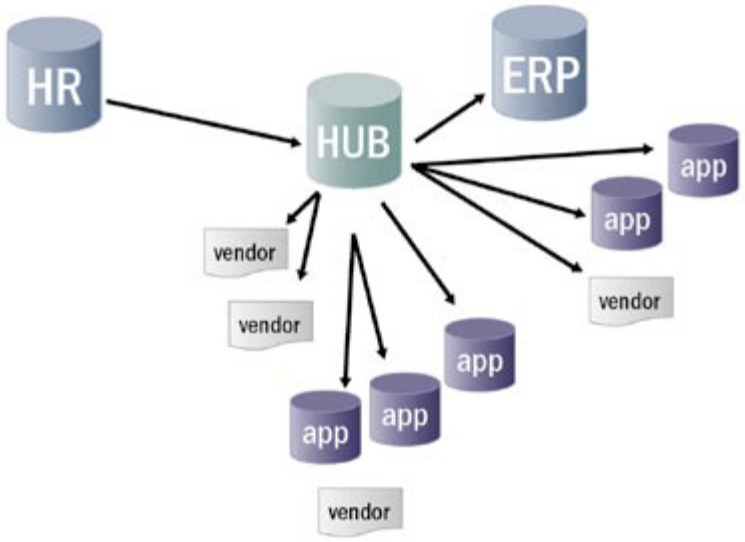


Figure 2. *Streamlined data hub built on PowerCenter saves development and maintenance costs and provides extensibility for future growth.*

The Results

Projected $1.2 Million Savings in Development and Maintenance Costs

Over three years, Seagate HRIT estimates it will save $1.2 million in development and maintenance costs previously spent on custom SQL coding across its complex and disparate integration touch points. It also expects significant cost savings through the retirement of legacy HR systems in the U.S. and Asia, reduction in change request backlogs, and a standardized integration environment that fosters shared skills among developers and helps eliminate reliance on specialized skills.

Enhanced Data Availability and Systems Performance

With PowerCenter orchestrating integration between the data hub and source and target applications, Seagate is exponentially improving the availability of data to the more than 100 target applications and their business users. With the legacy system, data was extracted from the Oracle HRMS application just once a day, and it took up to 48 hours to reach the downstream targets. PowerCenter now executes four data transfers per day from Oracle HRMS to the data hub, each in less than 10 minutes. And it delivers data from the hub to downstream targets between one and four times a day (according to business needs), so that information is updated in as little as two hours. For instance, it makes a stock plan application available to new employees within six hours of their hire.

Improved Data Consistency and HR System Efficiency

In the past, the five localized data stores that routed information from the Oracle HRMS application applied varying transformations before distributing the data to downstream applications. The result was differing data semantics and definitions across the downstream network, with discrepancies in headcount and other problematic issues. The PowerCenter-driven data hub eliminates the problem by applying a common set of transformations to all HR data before distribution and helps ensure the efficiency and accuracy of HR systems, as well as workforce productivity.

Ensured Extensibility for Future Growth and Changing Needs

The flexibility inherent in PowerCenter positions Seagate to rapidly respond to changing IT and business demands and fully leverage its global development resources. For instance, the central HR data hub may be readily extended to accommodate new business applications or additional outsource vendors. ■

LoanPerformance Revitalizes Competitive Web-Based Analytics Service

Commentary by David Gussmann, Sr. VP Product Development & Operations, LoanPerformance

Executive Summary

LoanPerformance supplies risk management and financial analysis tools to industry giants like Fannie Mae, Freddie Mac, Bank of America, JPMorgan Chase, Wells Fargo, and Washington Mutual. LoanPerformance's TrueStandings solution is a Web-based analytics solution that provides securities data to its customers. With Sybase IQ processing huge volumes of data, on-demand reports and queries are now delivered an average of 8 times—and up to 100 times—faster than before.

Customer Profile

LoanPerformance is a subsidiary of First American Corporation, a *Fortune* 500 company with more than 30,000 employees and revenues in excess of $6.5 billion. First American manages the largest property and ownership database in the U.S., covering 94 percent of the U.S. population and 97 percent of all mortgage transactions. A self-standing subsidiary, LoanPerformance provides state-of-the-art technology predictive analytics and reporting capabilities to the world's leading financial institutions that invest or trade in mortgage risk.

Business Challenge

An innovator in mortgage securities data, LoanPerformance was faced with growing their business and supplying solutions to increasingly demanding customer requirements while operating an efficient technology architecture. Two additional drivers were adding new Web-based capabilities and improving the existing functionality of analytic applications. As the central databases grew exponentially into the terabyte range, queries and reports were taking longer than anticipated, and customers grew weary of extended wait times for their information. One in 5 queries took longer than five minutes, and one in 20 queries would time out. In short, the IT team knew it was time to investigate alternatives beyond a traditional OLTP database on which to run their mission-critical analytics applications.

Selection Criteria

Knowing the future success of the company would ride on its ability to deliver fast, accurate, Web-based reporting, the IT team developed their standards for the new solution. Simple queries must be delivered in under a minute, and complex queries in under 20 minutes. They also required enhanced attribute capabilities, faster load times, and scalability for thousands of concurrent users, as well as anticipated future data growth, both in fields and in record number.

Why Sybase?

During exhaustive head-to-head testing, Sybase IQ emerged as the leader in price/performance. Even with increasing numbers of concurrent users, speed requirements were exceeded—at a lower cost than the competition. Other solutions, while meeting the baseline requirements, slowed dramatically when additional concurrent users were added. LoanPerformance selected Sybase IQ to allow the company to pursue their healthy growth plans while providing astounding performance on a low-cost Linux platform.

Results

Installation of the new Sybase IQ system took only one week and delivered dramatic results immediately. Queries taking more than 10 minutes were reduced from 16 percent to less than 1 percent. LoanPerformance also saw an average speed increase of 8 times—and up to 100 times—over the previous system, resulting in better customer service and better utilization of IT resources. Now LoanPerformance can continue to offer unparalleled Web-based analytics services and information from the industry's largest, most comprehensive database at lightning-fast speeds. So now, when an investor asks, "How does my mortgage portfolio compare over time with the market?" LoanPerformance can provide the answers. Online. Faster than they ever thought possible. ■

LOANPERFORMANCE
A Subsidiary of First American Real Estate Solutions™

Overview

Industry
Financial Services

Sybase Technology
Sybase IQ

Key Benefits

- Increased analytics speed by an average of eight times
- Queries taking more than 10 minutes reduced from 16 percent to less than 1 percent
- Rapid implementation
- Deployed on cost-effective Linux® platform
- Scalable for users and data loads

Business Intelligence Solutions Make Overstock.com a Smarter Company

Commentary by Jack Garzella, Vice President of IT Operations, Data Warehousing, Reporting, & Analytics, Overstock.com, Inc.

The Challenge

Founded in 1999 as online shopping was skyrocketing, Overstock.com, Inc., is an online outlet retailer offering discount, brand-name products to consumers across the U.S. The company realized that the Internet couldn't compete effectively with normal retail sales channels, but was tailor-made for product liquidation. Since then, Overstock.com has quickly become the online leader in a market valued at $60 billion.

Keeping track of inventory and sales on more than 650,000 products located in several warehouses—both internal and from partner factories—across the U.S. was a challenging task for the growing company.

With all the data being generated, Overstock.com knew they needed a new data warehouse system. They worked with Teradata to implement their data warehousing platform and then looked for a reporting tool that would serve as the front end, allowing business users to more easily view product inventory, decide which items to delete or expand to better manage inventory carrying costs, and much more.

"Our previous reporting system basically broke down because it could not keep up with the growth of the company. We had to turn it off to keep up with Web site volume," says Jack Garzella, who runs Overstock.com's IT operations and data warehousing. "And we had stale information, which wasn't very useful in making sound business decisions."

Overstock.com's reports were written in custom code against the operational systems and proved to be too difficult for the company to maintain. The company needed real-time information in an easy-to-read, understandable format.

The Approach

Overstock.com chose to standardize on a business intelligence (BI) platform from Business Objects because it easily and seamlessly integrated with their Teradata data warehouse. In addition, the company was already using Crystal Reports® from Business Objects throughout the organization.

Overstock.com's BI platform is made up of BusinessObjects™ Enterprise, BusinessObjects Web Intelligence™, Crystal Reports, Data Integrator, and BusinessObjects Rapid Marts™ (including the Accounts Receivable Rapid Mart, Accounts Payable Rapid Mart, General Ledger Rapid Mart, Purchasing Rapid Mart, and Sales Rapid Mart).

Currently, Overstock.com is using Data Integrator and Rapid Marts to pull financial information from its Oracle system and integrate it into the Teradata database. According to Garzella, about 120 to 130 tables—or a few million records—are transferred each night using Data Integrator.

Further, Overstock.com deployed BusinessObjects Enterprise to more than 200 business users across the company in marketing, merchandising, and the executive staff. "Basically any business person in the company can access the reporting portal and view reports," says Garzella.

The Results

Overstock.com is now better equipped to make critical business decisions. In addition to the Data Integrator and BusinessObjects Enterprise implementations, the company has also created dashboards with Web Intelligence that give employees an at-a-glance view of important company data such as daily sales history, item status, warehouse inventory, and more. The dashboards track revenue, gross margins, advertising expenses, how many consumers visited the site, and how many products were sold each day. Employees can also drill down into the information for more insight. "We can now give our employees real-time sales and shipping data all the way down to the SKU level," adds Garzella. "Overall, our BI solution is making us a much smarter company.

"The biggest benefit so far is the ability to see cause and effect. When we change a marketing campaign or a price, we are able to see the effect on sales or the effect on customer service and immediately take action," concludes Garzella.

Overstock.com is planning to implement new dashboards from BusinessObjects Performance Manager to help them monitor key performance indicators and provide more ad hoc capabilities for business users. ■

Overstock.com®
Your Online Outlet™

Industry: Retail

Business Pain

With the fast growth of the company, Overstock.com's old reporting system couldn't keep up. After implementing a new data warehouse system from Teradata, the company decided to find a new reporting tool that could provide real-time sales and marketing data in easy-to-read reports to functional role end users.

Why Business Objects?

With BI from Business Objects, Overstock.com now has real-time access to critical sales, inventory, or marketing information. Every business person in the company has access to view these reports—enabling Overstock.com's staff to make better-informed decisions.

Business Objects Products and Services

- BusinessObjects Enterprise
- Crystal Reports
- Web Intelligence
- Data Integrator
- Rapid Marts

Data Integration: Using ETL, EAI, and EII Tools to Create an Integrated Enterprise

By Colin White, President, BI Research

This report is a sequel to TDWI's 2003 report *Evaluating ETL and Data Integration Platforms*. The objective of the present report is to look at how data integration techniques, technologies, applications, and products have evolved since the 2003 report was published. The focus this time is not only on the role of data integration in data warehousing projects, but also on developing an enterprisewide data integration strategy.

The Challenges of Data Integration

Integrating disparate data has always been a difficult task, and given the data explosion occurring in most organizations, this task is not getting any easier. Over 69% of respondents to our survey rated data integration issues as either a *very high* or *high* inhibitor to implementing new applications. The three main data integration issues (see Figure 1) listed by respondents were data quality and security, lack of a business case and inadequate funding, and a poor data integration infrastructure.

Characteristics of Data Integration

Data integration involves a framework of applications, techniques, technologies, and products for providing a unified and consistent view of enterprise business data (see Figure 2).

- *Applications* are custom-built and vendor-developed solutions that utilize one or more data integration products.
- *Products* are off-the-shelf commercial solutions that support one or more data integration technologies.
- *Technologies* implement one or more data integration techniques.
- *Techniques* are technology-independent approaches for doing data integration.

Top Data Integration Issues

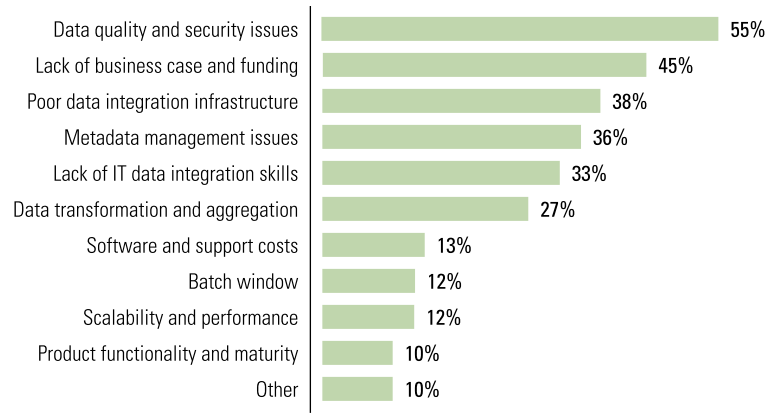


Figure 1. *The top inhibitors to the success of data integration projects. Respondents were asked to select up to three. Based on 672 respondents.*

Data Integration Framework

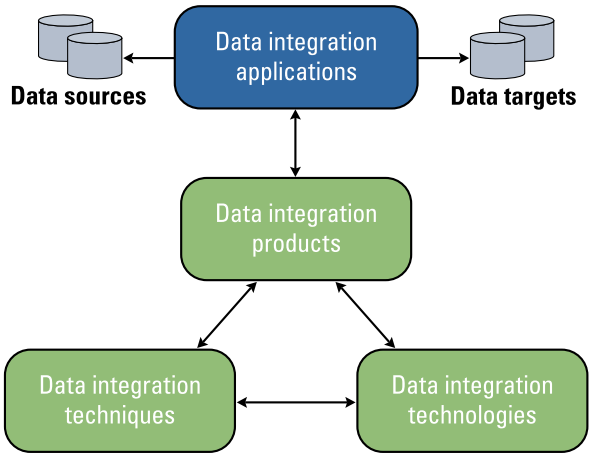


Figure 2. *Components of a data integration solution.*

Following is a review of the the techniques and technologies used in data integration projects.

Data Integration Techniques

There are three main techniques used for integrating data: consolidation, federation, and propagation.

Data Consolidation captures data from multiple source systems and integrates it into a single persistent data store. This data store may be used for reporting and analysis as in data warehousing, or it can act as a source of data for downstream applications as in an operational data store.

With data consolidation, there is usually a delay, or *latency*, between the time updates occur in source systems and the time those updates appear in the target store. Depending on business needs, this latency may be a few seconds, several hours, or many days. The term *near real time* is often used to describe target data that has a low latency of a few seconds, minutes, or hours. Data with zero latency is known as *real-time* data, but this is difficult to achieve using data consolidation.

Data Federation provides a single virtual view of one or more source data files. When a business application issues a query against this virtual view, a data federation engine retrieves data from the appropriate source data stores, integrates it to match the virtual view and query definition, and sends the results to the requesting business application. By definition, data federation always *pulls* data from source systems on an on-demand basis. Any required data transformation is done as the data is retrieved from the source data files. Enterprise information integration (EII) is an example of a technology that supports a federated approach to data integration.

Data Propagation applications copy data from one location to another. These applications usually operate online and *push* data to the target location; i.e., they are event-driven. Updates to a source system may be propagated asynchronously or synchronously to the target system. Synchronous propagation requires that updates to both source and target systems occur in the same physical transaction. Regardless of the type of synchronization used, propagation guarantees the delivery of the data to the target. This guarantee is a key distinguishing feature of data propagation. Most synchronous data propagation technologies support a two-way exchange of data between a data source and a data target. Enterprise application integration (EAI) and enterprise data replication (EDR) are examples of technologies that support data propagation.

A Hybrid Approach. The techniques used by data integration applications will depend on both business and technology requirements. It is quite common for a data integration application to use a *hybrid* approach that involves several data integration techniques.

Data Integration Technologies

A wide range of technologies are available for implementing the data integration techniques outlined above. This section reviews four of the main ones: extract, transform, and load (ETL); enterprise information integration (EII); enterprise application integration (EAI); and enterprise data replication (EDR). Master data management (MDM) and customer data integration (CDI), which are really data integration applications, are also discussed because they are often thought of as data integration technologies.

Extract, Transform, and Load

As the name implies, ETL technology extracts data from source systems, transforms it to satisfy business requirements, and loads the results into a target destination. Sources and targets are usually databases and files, but they can also be other types of data stores such as a message queue. ETL supports a consolidation approach to data integration.

Data can be extracted in schedule-driven pull mode or event-driven push mode. Both modes can take advantage of changed data capture. Pull mode operation supports data consolidation and is typically done in batch. Push mode operation is done online by propagating data changes to the target data store.

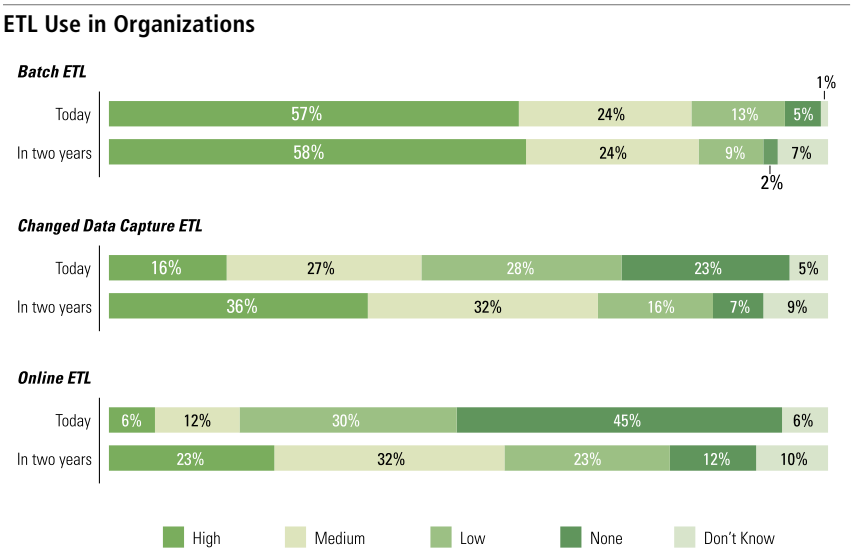


Figure 3. *Batch ETL use is flat, but changed data capture and online ETL use will grow over the next two years. Based on 672 respondents.*

Data transformation may involve data record restructuring and reconciliation, data content cleansing, and/or data content aggregation. Data loading may cause a complete refresh of a target data store or may be done by updating the target destination. Interfaces used here include de facto standards like ODBC, JBDC, JMS, for example, or native database and application interfaces.

In our survey, 57% of respondents rated their batch ETL usage as *high* (see Figure 3). Adding a *medium* rating to the result increases the figure to 81%. The survey also asked what the likely usage of batch ETL will be in two years. The result was 58% for high usage, and 82% for high and medium. As expected, these figures demonstrate that the batch ETL market has flattened out because most organizations use it.

The picture changes when looking at the growth figures for changed data capture (CDC) and online ETL operations. Our survey shows 16% of respondents rated their usage of CDC in ETL today as *high*. This number grows to 36% in two years. The equivalent figures for online ETL (called real-time or trickle-feed ETL in the survey) were 6% and 23%, respectively. These growth trends are due primarily to shrinking batch windows and the increasing need for low-latency data. It is interesting to note that combining the high and medium usage figures for the two-year projection of online ETL gives a result of 55%. This clearly shows the industry is moving from batch to online ETL usage.

Enterprise Information Integration

EII provides a virtual business view of dispersed data. This view can be used for demand-driven query access to operational business transaction data, a data warehouse, and/or unstructured information. EII supports a data federation approach to data integration.

The objective of EII is to enable applications to see dispersed data as though it resided in a single database. EII shields applications from the complexities of retrieving data from multiple locations, where the data may differ in semantics and formats, and may employ different data interfaces.

Distinguishing features to look for when evaluating EII products include the data sources and targets supported (including Web services and unstructured data), transformation capabilities, metadata management, source data update capabilities, authentication and security options, performance, and caching.

In our survey, 5% of respondents rated their EII use as *high* (see Figure 4). Adding a *medium* rating to the result increases the figure to 19%. These figures grow to 22% and 52% respectively in two years, indicating considerable interest in exploiting EII technology in the future.

Enterprise Application Integration

EAI integrates application systems by allowing them to communicate and exchange business transactions, messages, and data with each other using standard interfaces. It enables applications to access data transparently without knowing its location or format. EAI is usually employed for real-time operational business transaction processing. It supports a data propagation approach to data integration.

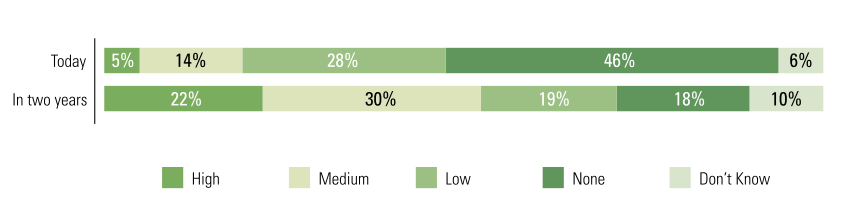


Figure 4. *EII use is low at present but its usage is likely to grow rapidly. Based on 672 respondents.*

The direction of the EAI industry is toward the use of an enterprise service bus (ESB) that supports the interconnection of legacy and packaged applications, and also Web services that form part of a service-oriented architecture (SOA).

From a data integration perspective, EAI can be used to transport data between applications and to route real-time event data to other data integration applications like an ETL process. Access to application sources and targets is done via Web services, Microsoft .NET interfaces, Java-related capabilities such as JMS, legacy application interfaces and adapters, etc.

EAI is designed to propagate small amounts of data from one application to another. This propagation can be synchronous or asynchronous, but is nearly always done within the scope of a single business transaction. In the case of asynchronous propagation, the business transaction may be broken down into multiple physical transactions. An example would be a travel request that is broken down in separate but coordinated airline, hotel, and car reservations.

Data transformation and metadata capabilities in an EAI system are focused toward simple transaction and message structures, and they cannot usually support the complex data structures handled by ETL products. In this regard, EAI does not compete with ETL.

In our survey, 9% of respondents rated their EAI usage as *high* (see Figure 5). Adding a *medium* rating increases the figure to 29%. These figures grow to 26% and 58% respectively in two years. It is important to point out that the question relates to the use of EAI for data integration, as opposed to the use of EAI in the organization overall. The two-year EAI projection of 58% is consistent with the 55% growth figure for online ETL use mentioned earlier. This suggests that organizations see the need to merge the event-driven benefits of EAI with the transformation and consolidation power of ETL.

EAI Use in Organizations

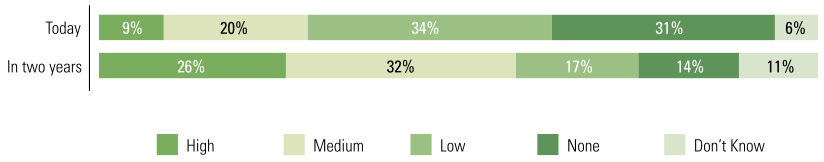


Figure 5. *EAI growth is consistent with the growth in online ETL use shown in Figure 3. This suggests the two technologies will be used together. Based on 672 respondents.*

Summary

This study looked at data integration approaches across a wide range of different companies and applications. The study results show that these companies fall into two main groups:

- *Large organizations* that are moving toward building an enterprisewide data integration architecture. These companies typically have a multitude of data stores and large amounts of legacy data. They focus on buying an integrated product set and are interested in leading-edge data integration technologies. These organizations also buy high-performance best-of-breed products that work in conjunction with mainline data integration products to handle the integration of large amounts of data. They are also more likely to have a data integration competency center.

- *Medium-sized* companies that are focused on data integration solely from a business intelligence viewpoint and who evaluate products from the perspective of how well they will integrate with the organization's BI tools and applications. These companies often have less legacy data, and are less interested in leading-edge approaches such as right-time data and Web services.

In evaluating and applying the contents of this report, it is important to understand which of the two categories your company fits into, and thus how sophisticated a data integration environment your company needs. Nonetheless, many of the ideas and concepts presented in this report apply equally to all companies, regardless of size. The main message of this report is that data integration problems are becoming a barrier to business success, and your company must have an enterprisewide data integration strategy if it is to overcome this barrier. ■

Colin White is the founder of BI Research. With over 35 years of IT experience, he has consulted for dozens of companies throughout the world and is a frequent speaker at leading IT events. Colin has written numerous articles on business intelligence and enterprise business integration, and publishes an expert channel and a blog on the Business Intelligence Network. He can be reached at info@bi-research.com.

This article was excerpted from the full November 2005 report. TDWI appreciates the sponsorship of Business Objects, Collaborative Consulting, DataFlux, DataMirror Corporation, IBM Information Integration Solutions, Informatica Corporation, SAP America, Sunopsis, and Syncsort Incorporated.

To download the full report, visit www.tdwi.org/research/reportseries.

Master Data Management: Getting the Most from an MDM Initiative

Matthew J. Beyer, Senior Architect, Collaborative Consulting

Without a doubt, master data management can help a company do business much more efficiently. When an entire organization and its partners understand the same information to have the same meaning, the company can make and sell products faster, connect more effectively with suppliers and customers, improve business process efficiency, and simply make more money. An MDM initiative also helps a company improve its data quality and consistency; develop applications more efficiently; and eliminate the costs and waste born of poorly managed data. But organizations are discovering that an enterprise MDM initiative can be a complex and expensive endeavor.

Cooperation and Sacrifice

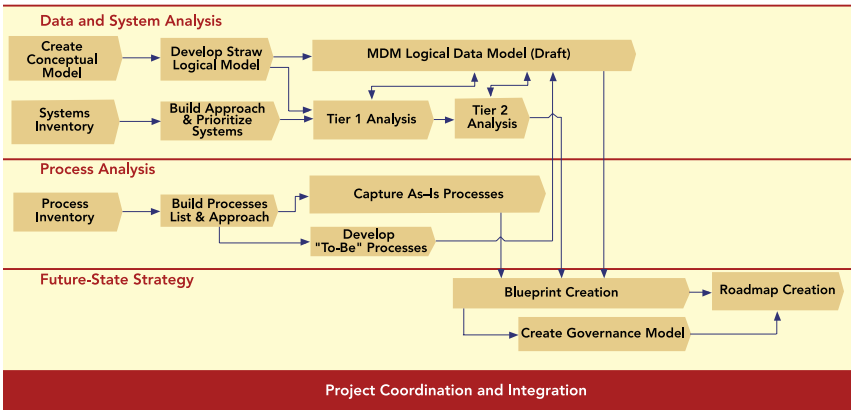
In fact, preparing for an MDM initiative requires substantial cooperation among numerous internal and external groups—a tall order for many organizations. To work optimally, people from several functions and departments must come to a consensus on master data—what comprises it, what defines it, and what its sources are. Some departments may have to make sacrifices and endure dramatic changes.

An MDM project's scope must be defined clearly to realize its promise. When teams maintain a sharp focus, gains made via MDM are often real and significant. Identifying business value–based focus areas is critical. Initiatives such as supplier and product "on-boarding," customer retention, and transportation optimization create an impetus for MDM as well as a tangible system and information scope. This focus, however, often means that not all parties benefit in the near term. As a result, prioritization and flexibility also become critical.

Executives considering an MDM initiative can learn a valuable lesson from organizations that have tried to tackle an enterprise data warehouse all at once. Many large-scale data warehouse projects include many, if not all, of a company's subject areas. This approach is risky because of the lengthy time frame involved and the extreme complexity in developing an entire solution on the first pass. A safer, more effective approach, one that assesses, prioritizes, and implements iteratively, has a much higher chance of success. This also holds true with MDM.

By assessing the existing environment, prioritizing areas that reap the greatest benefits, and implementing in a phased manner, success is considerably more likely than with a "big-bang" method.

MDM Initiative



Multi-Phase Approach

Once the strategy has been defined, a practical, multi-phase approach must be developed. The first phase (technology and data management) entails assessment and analysis of systems and applications with regard to master data—where data is stored and how it is managed or used. It also includes activities that define master data and develop a logical data model.

The second phase concerns business process and data quality—documenting existing processes and developing future ones to support the solution. Master data management is as much about process improvement as technology implementation. A data governance phase follows. Many projects/systems falter due to a lack of ongoing sponsorship and commitment. What once seemed like the right solution, if not constantly revisited and updated, may quickly become inappropriate in a rapidly shifting business climate. It is especially critical to identify definitive data and data management process owners to create the appropriate environment for clear data definition, consistent data management, and accountability for data availability and quality. Establishing a governance model that will work within the targeted organization is integral to long-term success.

Keep it Real

Finally, companies need to employ a solid dose of realism. Often, MDM is not about centralizing data into a single tool, but centralized management of data across the existing application set. It may be more prudent to extend the capabilities of the existing applications and improve data integration than to try to implement a centralized MDM toolset. Sensible, pragmatic expectations are critical; in fact, the companies that have the most success with MDM are honest with themselves, employ a practical approach, and maintain a holistic, open-minded perspective. ■

BI Be Nimble, BI Be Quick

Efficient and Long-Term Plans for Business Intelligence

By Megan Lordeon, Director of Strategic Marketing, MicroStrategy

Leading companies are under pressure to select and deploy business intelligence solutions that are both efficient and affordable for the long term. Seasoned IT leaders realize that business intelligence is a permanent strategy; therefore, they plan for the long term by selecting software and hardware that allows them to grow affordably and efficiently. Many have tried homegrown or cobbled-together solutions that eventually began to drain corporate resources, as they were expensive to maintain and time-consuming to administrate.

The new era of BI promotes the efficient use of human, network, and hardware resources. Companies with this type of BI solution can empower business users with interactive reporting, reduced report creation time, seamless reporting and analysis, and a reduced number of reports to design and maintain.

Empower Business Users

Interactive reporting enables business users to create and tune their reports without assistance from already overworked IT staff or BI administrators. Companies should select a BI platform that allows users to interact fully with their reports in a what-you-see-is-what-you-get (WYSIWYG) manner over the Internet. Business users instantly see the effects of any changes they make in reports that are fully populated with live data, making report editing easier.

The WYSIWYG design delivers Windows-like functionality over the Web, giving business users a familiar and intuitive reporting environment. They can undo and redo their design ideas as often as necessary to refine the format without requiring assistance from IT developers. Thus, companies gain efficiency in business user empowerment and self-service.

Reduce Report Creation Time

The WYSIWYG design functionality radically reduces the time it takes businesses to deploy new reports. Without a BI platform that has this design feature, report developers are required to develop the entire report through seemingly endless iterations with business users. Companies can eliminate the need for these onerous back-and-forth iterations by selecting a platform with WYSIWYG functionality.

Without assistance from IT developers, business users can take ownership of the final steps in the design process by making their own finishing touches to the report—changing report layout and format by repositioning data and sections; adding or removing borders and background colors; changing fonts, headers, and labels; graphing data; and even adding new calculations. By streamlining report development, business users are far more efficient and IT staff are more effective.

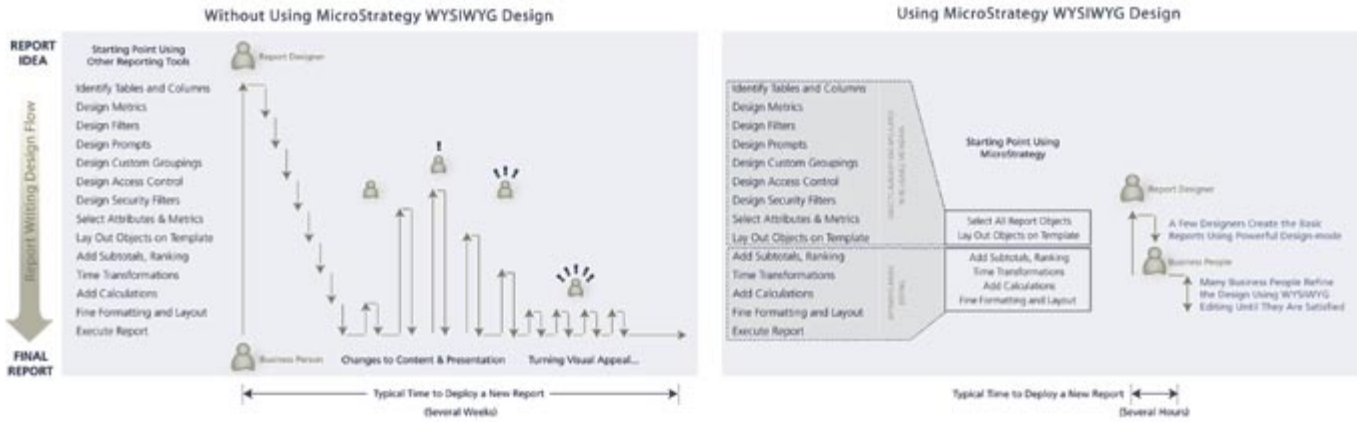
Seamless Reporting & Analysis

More than ever, business users are crossing the line from reporting to analysis, and from consuming data to investigating it. Select a BI architecture that offers the ability to integrate both reporting and analysis, thereby truly meeting business users' needs for information without boundaries.

Reduce the Number of Reports Designed and Maintained

Business users should have the ability to specify what content to include in a report each time they run it by answering intuitive prompts. A single report design should be able to produce hundreds of variations through a parameterized reporting capability.

MicroStrategy ushers in a new era of business intelligence by delivering the most resource-efficient platform, built for affordable, long-term BI strategy. MicroStrategy has revolutionized interactive reporting by streamlining report development for business users and IT staff, reducing the number of reports designed and maintained, and delivering more analytical insight in each report. ■



Appliances—Data Mart or Enterprise Data Warehouse?

By Stuart Frost, CEO, DATAllegro

Appliances are becoming established in the data warehousing market, but some companies and analysts have positioned appliances as "just" suitable for data marts (DM). Is this true, or can they also be used for large-scale enterprise data warehouse (EDW) projects?

The answer is yes, they can—under certain circumstances. While few would claim that appliances are currently ready to handle complex EDW, appliances are finding an interesting niche as an integral part of many EDW infrastructures.

DM and EDW Differences

Definitions of DMs and EDWs vary, but the most common differences lie in the number of business processes supported by a given system. A DM typically supports only one business process or subject area, whereas an EDW supports several, and in some cases is a true enterprisewide system. In addition, DMs are often fed summarized information from the EDW in a hub-and-spoke architecture, although this varies across the industry.

> Since appliances are relatively easy and cheap to maintain, any additional complexity … is limited in nature and overwhelmed by the huge benefits.

EDW Challenges

A significant majority of Global 2000 companies have deployed data warehouses in the last 10 years, establishing the overall business value of analytics. However, many companies are now struggling to keep up with new demands on their data warehouse systems. Such challenges include:

- Significant data growth due to:
  - New legislation (the Sarbanes-Oxley Act, EU data retention laws, etc.)
  - Mergers and acquisitions
  - The need to analyze growing volumes of point-of-sale or telecommunications transactions to remain competitive
- Business demands for reduced latency, which translates into faster query times
- Larger user bases
- Demand for ever more complex, ad hoc queries to address fraud detection and anti-money laundering

As a result, many previously successful EDW installations on platforms such as Teradata, DB2, and Oracle are becoming overwhelmed by the need to support hundreds of users with a broad mix of query types against tens of terabytes of data. Upgrade quotes for these platforms can easily be tens of millions of dollars—and even then they may not meet business needs!

Using Appliances to Divide and Conquer the Problem

Since high-performance data warehouse appliances are now available at prices as low as $20,000 per terabyte, a number of EDW users are turning to this new technology as a potential solution. However, they are not relegating appliances to the role of mere data marts. Instead, they are using appliances as a low-cost front end to the EDW itself.

In a typical scenario, large-volume, fine-granularity transaction records are stored directly on the appliance. The appliance then handles tasks such as:

- Data cleansing
- Long-term storage of transaction details for compliance
- Ad hoc queries
- Applications such as fraud detection that require access to data at very fine granularity
- Exports to external analytics systems such as SAS
- Building large-scale aggregation or summary tables and exporting them to the EDW

By offloading these tasks from the EDW to the appliance, companies are greatly reducing the need for an expensive EDW upgrade. In addition, the specialized nature and advanced technology of the appliance enables these processes to run significantly faster, often by two orders of magnitude.

Since appliances are relatively easy and cheap to maintain, any additional complexity introduced by this divide-and-conquer approach is limited in nature and overwhelmed by the huge benefits.

Summary

New data warehouse appliance technologies have the potential to transform the data warehousing market. By acting as a high-performance, high-capacity, and low-cost front end to an established EDW, they can add significant value to an already successful installation—while avoiding expensive upgrades.

If this all sounds too good to be true, many vendors offer free proofs of concept so you can check out their claims at minimal cost. What do you have to lose, apart from poor performance and high costs? ■

The W5 of Enterprise Information Management

By Darren Cunningham, Product Marketing Director, Business Objects

Enterprise information management (EIM) includes the strategy, practices, and technologies that allow organizations to manage disparate data effectively. The right approach to EIM will help drive individual and organizational performance, enable business intelligence (BI) standardization, and ensure that relevant and timely information is delivered to business users in a way they understand. But when you think of EIM, it's important not to think of it just as middleware. Instead, you need to consider "middle-where," "middle-why," "middle-what," "middle-when," and "middle-who?"

Middle-Why?

If your goal is to become a high-performance organization, where people are connected to common metrics, goals, and strategies, you must not underestimate the importance of the underlying data foundation. Is your organization seeking to attract and retain skilled workers? Are you hoping to acquire new customers? Is your company aiming to increase shareholder value while continually adapting to rapidly changing market conditions? Do you want to use IT to reduce costs and create more value?

For most organizations, the answer would be a resounding, "Yes!" The work done behind the scenes—the work that end users never see—will be the key to this success.

Middle-Where?

Determining what your business users require and what tools they're using today is critical to an EIM strategy. A tools-and-skills audit will give you the current picture and allow you to develop a long-term vision and plan on how to reach your BI goals. Knowing where to focus first (department, geography, level, etc.) will also help you get some early wins and establish credibility as you roll out initiatives like performance management dashboards and scorecards. Focus on the bottom (systems and data), top (users), and middle (data quality and integration).

Middle-What?

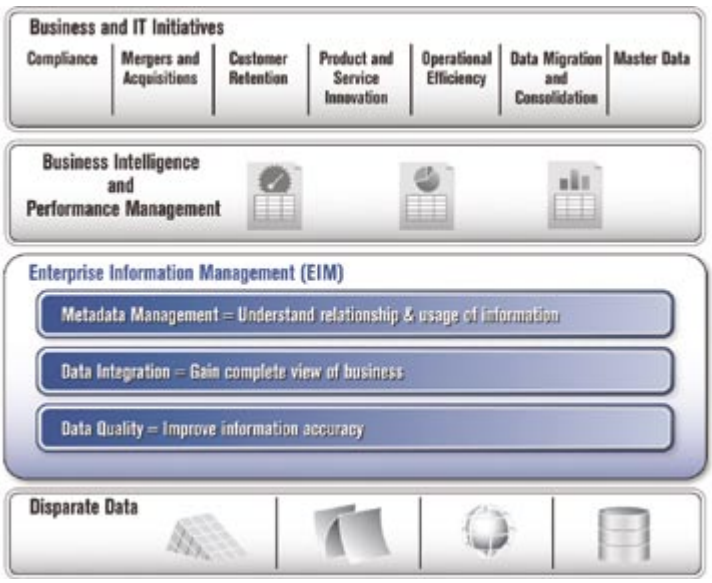
Whether it's a data warehouse or data store, there are many tools available to help you physically or virtually move data and make it relevant to BI users. ETL tools are used to move large volumes of data in batch from multiple sources to targets, and EII tools enable federated querying capabilities across systems. Either way, the right EIM strategy helps to solve data centralization, reconciliation, and efficiency issues. Without data quality, there is no end-user confidence or trust in the numbers. Many organizations are now creating the role of the data steward to help deliver what essentially becomes a "data quality firewall." Master data management (MDM) practices use technologies to provide a layer of standard definitions about the business (customer, product, sales). Some organizations are implementing MDM for data quality purposes, helping them to transform data to a common standard.

Middle-When?

When it comes to EIM, don't wait to get started. That does not mean you should simply pour more resources into new tools. An EIM strategy that is directly tied to your BI and performance management initiatives will begin with the end in mind and enable you to architect a solution based on the unique information access and analysis requirements of your entire BI user community.

Middle-Who?

The key to the right infrastructure and a comprehensive EIM strategy is both data neutrality and data proximity. Make sure your vendors are able to take advantage of your transactional systems, but are not inherently tied to them in terms of key features and benefits. From a BI, data-integration, and data-quality perspective, it is critical to choose components that are tightly integrated. Few vendors can claim to deliver everything in an integrated way, so ask them about the depth and breadth of their product and services offerings, and seek to establish the elusive "trusted advisor" relationship with a select few. ■



Enterprise information management: *Strategy, practices, and technologies to effectively manage disparate data and drive performance.*

The Managed Spreadsheet Environment

How to Preserve the Best and Most Useful Features of Spreadsheets

By Neil Raden, Founder of Hired Brains, on behalf of Actuate

It's commonly assumed that spreadsheets became widely deployed in organizations because they were intuitive, easy to use, and filled a need that was not being met by other technologies. The truth is that spreadsheets exploded in popularity because they were used as an alternative to enterprisewide, IT-backed reporting, planning, and analysis applications.

Today, spreadsheets are used for complex multi-user application development, stretching to the breaking point their already razor-thin capacity for security, maintainability, scale, networking, and concurrent use. Because of changes in the regulatory environment, such as the Sarbanes-Oxley Act, these applications, no matter how well conceived and executed, can potentially expose organizations and their executives to fines and even criminal charges if they lack sufficient transparency.

As a prototyping tool, spreadsheets have no equal, but with constrained IT budgets (the norm over the past few years), an attractive prototype can become a permanent stand-in. The drawback is that the feasibility curve for spreadsheets hits the wall as the complexity increases, and the applications either fail precipitously or gobble up greatly underestimated maintenance costs to keep them functioning.

To make matters worse, a large part of these maintenance costs is the time of manager-level people. This not only increases costs in an obscured way, but also diverts the attention of critical people from their primary roles. The principal contributors to complexity are scale (number of users, amount of data), serial models (linked spreadsheets, dependent applications), external data such as customer lists or syndicated data, rapidly changing elements, assumptions, and organizational structure.

At the extreme, spreadsheet applications attempt to perform operations that are clearly best suited for enterprise tools, such as data cleansing and integration, and often employ programming with scripting or Visual Basic, which places them squarely in the realm of software engineering.

Not all spreadsheet applications suffer from dangerous deficiencies in security, scalability, error trapping, etc. There are many instances of clever, even brilliant, workarounds that stretch spreadsheet software far beyond its intended purpose. The problem is that, in most cases, the workarounds are *not* brilliant, and they generously substitute labor for good design.

The use of spreadsheets is a crucial part of virtually every organization's information management process, and they contribute a great deal of value. Banning them or trying to dictate their use has proven to be counterproductive.

Today, spreadsheets are used for complex multi-user application development, stretching to the breaking point their already razor-thin capacity for security, maintainability, scale, networking, and concurrent use.

ACTUATE®

How are spreadsheets used productively? What roles do they perform that are vital to an organization? Typically, spreadsheets are employed to fulfill roles in three primary activities:

1. Gathering, creating, and managing data
2. Creating models and calculations
3. Disseminating information, principally as reports

Gathering, creating, and managing data represents an entire discipline within IT, so the thought of domain experts with no formal training in data management devising applications is worrisome to most IT managers. Models and calculations, on the other hand, are more or less alien to mainstream IT, which is more focused on infrastructure and transaction processing.

Nevertheless, a derived value shares equal weight with source data, and modeling processes must have a formal review and approval process, especially today. Only the last activity, disseminating information, is relatively risk free. The other two are the principal source of error and cost. A truly breakthrough solution would be one that preserves all of the features of current spreadsheets for personal use, but manages the use of spreadsheets for gathering data and creating models and calculations for shared applications.

There are alternatives to the Shadow IT conundrum. It is not necessary to throw away the baby with the bathwater. All that is needed is to unbundle the best parts of spreadsheet programs from the practices that lead to dysfunction, excessive costs, and risk in order to provide better-managed solutions that suit everyone's goals. ■

Data Quality and the Emergence of Customer Data Integration

By Tony Fisher, President and General Manager, DataFlux Corporation

Several years ago, a large manufacturing company lost a key distribution center to a fire. The fire destroyed not only the building, but also thousands of shipments destined for a global customer base.

Naturally, the CEO of this company wanted to send a letter to customers to explain the situation and to provide a timetable of when operations would return to "business as usual." The CEO passed the request to the vice president of customer relations, who in turn asked IT to generate a list of all customers for that particular center.

The IT staff pulled reports from its enterprise resource planning (ERP), customer relationship management (CRM), billing, and supply chain management systems. What they found reveals a distressing but pervasive problem at most organizations: each list contained different, overlapping, and confusing "views" of the customer base. The end result was that this company could not create an accurate and inclusive list of the customers affected by the loss of the distribution center.

For years, enterprise applications have been promising a single view of the customer. However, the proliferation of systems has led to more confusion—at a data level—about the customer base. In fact, marketing and customer relations executives are struggling to understand even the most basic questions: Who exactly are our customers? Which customers are we trying to target? Who are our best customers? Which customers represent our best opportunities?

Uncertainty about customers' identities can severely compromise efforts to build stronger relationships. And in today's competitive marketplace, if customers don't feel valued, they will take their business elsewhere. Customers are hard to acquire, but even harder to keep.

Adding Data Quality and Identity Management to the Customer Equation

Customer data integration (CDI) is an emerging method for compiling the most authentic customer information from all applications, databases, and customer touch points into one centralized data source. By bringing the best information about customers to the surface, CDI strives to deliver consistent, accurate, and reliable information—regardless of the originating application.

CDI solutions are helping companies create consistent, accurate, and reliable data—and deliver a truly unified view of their customers that builds a firm foundation for sales, support, and marketing functions.

The benefit is that the data itself—not the applications—is the focus. Each business unit can view the same information about customers, which improves support and service across business functions.

Companies are now turning to CDI solutions with two added components: robust data quality capabilities and sophisticated identity logic (identity management). With these components, users can improve the quality of data while also identifying and managing the same customer sets across sources and applications.

The data quality component typically begins with an in-depth data profiling phase. The company then builds in business rules to standardize and verify addresses and other attributes, reconcile conflicting information, validate name and address information, and add demographic data to enhance the value of information.

The second component is identity logic, a crucial phase of any successful CDI effort. This determines whether customers listed in different sources are indeed the same customer, and intelligently integrates customer information from multiple applications and databases. The various records for Michael William Smith, Mike Smith, and Michael W. Smith, for example, are determined by identity logic to indeed be the same individual, provided other data points are similar. Companies can flag information for linking customers across applications and sources, and isolate the best data from multiple sources.

Building Lasting Customer Relationships

CDI solutions are helping companies create consistent, accurate, and reliable data—and deliver a truly unified view of their customers that builds a firm foundation for sales, support, and marketing functions. Thanks to CDI, organizations are developing healthier, more lasting relationships with their customer bases and can market more intelligently—and profitably—to these customers. ■

Data Mart Consolidation

By the Redbook Team, IBM Corporation

Just as many people today undertake a personal fitness program in order to reduce weight, improve health, and build stamina, a growing number of large organizations are taking a similar approach to their existing business intelligence solutions.

It is common for organizations that have used business intelligence systems over several years to find that they have a collection of different systems serving a similar purpose. These business intelligence systems are also referred to as management information systems (MIS) or data marts. While the reasons for this proliferation are natural and understandable, it is clear that it can cause a number of problems:

1. Multiple MIS systems make it difficult to be sure that you are using the right numbers when reporting internally and externally. There are many different versions of the truth. This is especially true when an organization uses many hundreds or thousands of spreadsheets to aid reporting.

2. The cost of running these multiple systems is much greater than the cost of a single consistent system.

3. Lack of productivity. The effort required in developing reports in different systems, getting the right report from the right system, and checking the reports for accuracy, is much greater than it needs to be.

The answer to this dilemma is to perform a data mart consolidation, where the different systems are taken and transformed into a single, consistent, shared system with comprehensive data of known quality. Streamlining and efficient tools can then be utilized to promote modern best practices.

The key benefits that can be obtained by undertaking such a program include:

1. Enhancing corporate agility by becoming even more responsive to new requirements and opportunities.

2. Improving governance and eliminating uncertainty by establishing a single, modern system.

3. Improving productivity through the application of current best practices and modern tools.

The answer to this dilemma is to perform a data mart consolidation, where the collection of different systems are taken and transformed into a single, consistent, shared system with comprehensive data of known quality.

4. Reestablishing data quality for confidence in decision making and regulatory reporting (important for compliance with Basel II, Sarbanes-Oxley, IAS, IFRS, etc.).

5. Securing large cost savings by eliminating redundant data and systems, thus cutting maintenance costs.

There are many understandable reasons why MIS systems or data marts proliferate—maybe an organization has a history of creating a new data mart or MIS system every time a department has a new analysis requirement. The result, however, is high support and maintenance costs for all the data marts that accumulate over time. These are usually funded independently at the department level, so the true overall cost to the organization may not be visible.

Similarly, organizations that grow by acquisition inherit the business intelligence systems of the acquired company. When a realistic picture of the existing marts is compiled and analyzed, the facts revealed can cause significant concern for CFOs, CIOs, and IT managers.

By taking action, the overall cost of consolidation typically pays for itself several times over—as well as improving business effectiveness and competitiveness. ■

Data Aggregation—Faster Performance, Better Business Management

By Rich Ghiossi, VP Product Management and Marketing, HyperRoll

Over the past 10 years, businesses have made enormous investments in systems designed to transform raw data into business intelligence. Yet as data volumes and the demand for information have grown, even the most powerful database servers have been unable to keep up. Performance problems have proliferated to the point where reports are taking hours or even days to run.

Adding to this problem is the changing face of the business intelligence user. Business intelligence has expanded from just a few power users performing strategic analysis to thousands of front-line managers and top executives who need operational updates as business events are happening. These people have increased the need for fast access to information just as business intelligence systems are slowing down.

The impact of slow system performance is far-reaching. Strategic business intelligence analysis requires iterations: to get the right answer, one must ask the right question. Unfortunately, the right question usually makes its appearance after many iterations, and slower performance means fewer iterations. As for making front-line decisions based on the latest data, having to wait hours or days for key reports can seriously impede operational analysis and delay even the simplest business decisions.

Data Aggregation: A New Alternative

To date, database administrators have had few alternatives for improving data performance. Most have tried to optimize the performance of the existing infrastructure either by "tuning" database structures or queries, or by building dozens of summary tables to anticipate user requirements. But because neither of these labor-intensive approaches truly addresses the root problem, performance improvements are nominal and always temporary.

Today, businesses are finding that the single most dramatic way to affect performance is through innovative **data aggregation techniques**, which can accelerate query performance by a factor of hundreds or even thousands. Data aggregation involves the consolidation of information into a highly summarized result.

The performance improvements achieved through these new techniques can impact multiple areas of the business, including:

- **Financial close.** Avoiding delays in the close due to slow revenue, balance sheet, and profit and loss reports helps companies meet their regulatory requirements.
- **Performance management.** The ability to generate ad hoc, near-real-time reports dramatically improves visibility into the health of the business.
- **Compliance.** High-performance reporting offers financial transparency throughout the organization, allowing decision makers to respond promptly to material events and resolve non-compliant activities quickly.
- **Supply chain.** Intraday access to information from partners, suppliers, employees, and customers enables operational performance management and enhances quality control.
- **Customer relationship management.** Intraday access to sales, utilization, and profitability numbers enables managers to react quickly to issues as they arise, not days later.

Data Aggregation from HyperRoll

HyperRoll's patented data aggregation software delivers immediate and substantial improvements in database performance. One HyperRoll financial customer was able to improve its query performance from 14 minutes to 9 seconds—increasing the query performance time by a factor of 93, and empowering analysts and decision makers to achieve rapid business insight.

HyperRoll works seamlessly within customers' existing relational reporting environment. HyperRoll is completely transparent to business intelligence users and their applications. Analysts and decision makers can continue to use the same tools, applications, and routines they have been using—only faster, better, and more effectively. ■

Impacted Business Area	Before HyperRoll	After HyperRoll	System Improvement
Financial close	18 hours	15 minutes	72 times
Performance management	30 hours	15 minutes	120 times
Compliance	40 hours	>2 hours	20 times
Supply chain	80 minutes	2 minutes	40 times
Customer relationship management (CRM)	60 minutes	5 seconds	720 times

HyperRoll Inc.

What circumstances prevent widespread adoption of operational BI?

Performance is one of the key requirements of operational BI—people making operational decisions need information in seconds, not hours or days. Unfortunately, most BI solutions can't deliver this level of performance, and that prevents companies from tapping the potential of operational BI. The good news is that as companies add data aggregation software to their reporting environments, we're now seeing operational BI become a reality. Data aggregation software from HyperRoll, for example, accelerates existing reporting solutions to deliver information in seconds. Faster performance will be the catalyst for widespread operational BI.

Analyst Viewpoint

Many managers assume that operational BI will kill the performance of their transaction systems. They continue to rely on the operational reports that they already have. Although these reports are usually static in nature, not timely, difficult to change, and may track the wrong metrics, organizations are reluctant to deploy flexible reporting and analysis tools on top of their operational data. However, the technology exists where operational BI can be safely deployed by implementing BI technology that has access not only to transaction system detail, but also to low-latency, aggregated data where managers and analysts can quickly recognize trends.

Request more information about HyperRoll Inc.

IBM Corporation

Do I need to choose one type of OLAP technology (ROLAP or MOLAP), or are there reasons to implement both? Also, is Linux ready for hard-core business intelligence?

Different users require different tools, so it may not be possible to standardize on just one. However, there are benefits to reducing the number of supported tools. Look into creating a business intelligence competency center to support your tools, and ensure your database supports many vendors' OLAP functions and metadata transfer.

And, yes! Linux is ready. Look into the various offerings from major vendors and determine which best meets your needs.

Analyst Viewpoint

The OLAP versus ROLAP argument boils down to speed of analysis versus flexibility. OLAP is usually the analyst's choice, and ROLAP favors reporting. Although evolving database technologies and client tools are blurring the lines between these two classes of technologies, performance differences continue to exist, and an organization that wishes to get the most out of its data resources will support both.

Although open source BI technology is still in the development stage, there are increasing numbers of mainstream BI and database vendors that have robust Linux offerings. Linux has become a key operating system in the BI server space.

Request more information about IBM Corporation

Informatica Corporation

We'd like to move to a more standardized approach to data integration across our organization, but how do we justify that to the business?

IT organizations can do three things to demonstrate the business value of enterprisewide data integration:

1. Tie data integration projects to specific, urgent business initiatives with measurable impact, such as merger and acquisition consolidation or regulatory compliance.
2. Avoid the "big-bang" approach. An enterprise data integration infrastructure should be rolled out incrementally to reduce risk.
3. Emphasize the need for robust data governance to ensure data quality, auditability, and availability, and to manage and protect information as a valued enterprise asset.

Analyst Viewpoint

Individual projects that involve data integration, such as migration to a new ERP or projects that involve data warehousing, are often financially justified as part of a package of larger benefits. The irony of this situation is that in order to maximize benefit-cost ratios, the selection of a data integration technology for a given project may exclude features needed on the next project. This leads to multiple data integration tools and strategies that, taken in the whole, are more expensive. Organizations that implement a data integration competency center with an enterprise-class data integration technology will reap the benefits of scalability.

Request more information about Informatica Corporation

MicroStrategy

How important is seamless reporting and analysis to my business users?

More than ever before, business users cross the line from reporting to analysis and from consuming data to investigating it. Seamless reporting and analysis empowers business users to glean more information from every report, all within one interface.

Seamless reporting and analysis is essential for business user workflow processes and the prudent use of IT report development resources. As business users are more empowered, the IT staff is no longer the bottleneck for new report development.

With true seamless reporting and analysis, companies gain administrative and resource efficiencies because they no longer have to deploy separate dashboarding, reporting, and analysis technologies. The most advanced BI platforms infuse every report, scorecard, and dashboard with the ability to fully analyze the underlying data.

Analyst Viewpoint

With a few exceptions, BI vendors are moving or have moved toward seamless reporting and analysis. They are responding to customers desiring to maximize BI benefit-cost ratios. Not only does this technology simplify implementation, training, and support costs, it also makes users more efficient, whether they are casual consumers of information or full-time analysts. Casual users who view reports but also want to do some limited exploring can do so without having to learn another tool. Analysts will be freed up from having to respond to simple information requests and will be able to spend their time on more complex issues.

Request more information about MicroStrategy

Oracle Corporation

What's database partitioning, and what can it do for me?

Database partitioning is when a database is divided into separate components (partitions) that are distinct, independent parts. It's a key tool for building and maintaining multi-terabyte databases, since it offers considerable manageability, performance, and availability advantages.

With partitioning, maintenance operations can be focused on particular portions of tables, allowing DBAs to pursue a divide-and-conquer approach. Maintenance can be performed on certain parts of the database while others remain up and running.

Performance can be improved through partitioning by limiting the amount of data to be examined or operated on. And from an availability perspective, if one partition of a partitioned table goes down, the other partitions will remain online and available to users.

Analyst Viewpoint

Database partitioning means getting the most for your database dollars. Data warehousing initiatives or mergers and acquisitions can put significant upward pressure on database storage. Although most modern database technology can store large amounts of data, providing high-access performance along with high availability are other issues. Partitioning plays a significant role in large database performance and maintenance. The leading database vendors have partitioning capabilities. However, not all take the same approach. Some products are geared toward transaction systems, and others toward data warehousing. Organizations will be well served by making sure that their database partitioning capabilities match their needs.

Request more information about Oracle Corporation

Sunopsis

What is E-LT, and how is it different from ETL?

E-LT stands for extract, then load and transform (as opposed to ETL, which means extract, transform, and load). The difference may seem subtle, but this swapping of letters makes a big difference with regard to architecture and performance. With the ETL approach (used by the majority of integration tools today), all of the data has to transit through an ETL engine. Worse, when a lookup of target data is required, data used for this lookup also needs to be moved from the target to the ETL engine, where all processing occurs. The E-LT approach, implemented mostly by new-generation integration software, uses the power of the RDBMS engine to execute all data mappings and transformations. All heavy-duty processing is done on the target (or in the sources when appropriate), leveraging the dataset-processing capabilities of RDBMS engines and decreasing data exchanges over the network. The performance gain of E-LT over ETL can reach several orders of magnitude!

Analyst Viewpoint

Vendor support for E-LT (extract, load, and transform) as opposed to ETL (or extract, transform, and load) recognizes that many data warehousing teams have done their transformation programming in the target database itself. Procedural extensions to database languages such as PL/SQL and TSQL have made these technologies quite robust in capability and performance. When a data warehousing program has limited development staff, the database administrator will likely be pressed into service for programming duties. Because it relies on the transformation capabilities of the database, E-LT technology should provide that database administrator with a greater comfort level and quicker start-up time.

Request more information about Sunopsis

Sybase, Inc.

Is it possible to accelerate the performance and reduce the complexity of my BI system while using the hardware and software my team already knows?

Some performance improvement can be gained through tuning and optimization techniques, technology upgrades, additional hardware, and hardware upgrades. But these methods are costly and have limitations. You might consider a preintegrated appliance, but these are based on proprietary, preconfigured platforms that can get expensive as your BI demands scale up.

A column-based analytics server is an economical choice to deliver the speed, scalability, and query flexibility required by users without requiring changes to your organization's BI ecosystem. And typically, a column-based analytics server has built-in features that reduce the complexity of managing the BI ecosystem—requiring fewer DBA cycles, using standard hardware and OS, and integrating easily with current BI tools and applications.

Analyst Viewpoint

Poor performance limits the usability of a business intelligence system. Good data access technology cannot overcome the limitations of a poorly designed and/or poorly tuned database. However, organizations seeking top performance from their business intelligence systems while reducing system complexity may want to consider specialized analytics servers. Unlike general-purpose database management systems that are designed to support transaction applications, these specialized servers are designed especially for business intelligence deployment. When choosing an analytic server, a wise organization will look for one that has a standard interface with their business intelligence technology. It should also require minimal maintenance.

Request more information about Sybase, Inc.

Syncsort Incorporated

We need to understand more about our customers' purchases, such as when they first purchased and the amount of their largest purchase. How can we quickly and easily compile this information?

Data warehouse experts agree that aggregates are the best way to speed warehouse queries. A query answered from base-level data can take hours and involve millions of data records and millions of calculations. With precalculated aggregates, the same query can be answered in seconds with just a few records and calculations. DMExpress provides several functions for aggregating your data. You might try creating an aggregate that keeps the dates of the first and largest purchases, the first purchase amount, the largest purchase amount, and the average and total purchase amounts. To rank your customers by number of orders, take a count of the records summarized.

Analyst Viewpoint

These questions and future queries about customer behavior can be answered from a well-designed database. Such a database will contain history (such as when the customer made his/her first purchase), and it will be integrated, since we are interested in purchases that might possibly involve data aggregations over different products and different locations. A savvy database designer will also include summary tables that precalculate data that can provide answers for questions (such as the total amount spent by customers on product *x*). A high-performance ETL tool will help produce these summaries with a minimum of delay.

Request more information about Syncsort

The Importance of Metadata for ETL

By Yves de Montcheuil, Director of Product Marketing, Sunopsis

According to Wikipedia.org, *metadata* (Greek *meta* + Latin *data*), is literally data about data: "information that describes another set of data." Metadata is arguably one of the cornerstones of data warehousing and all its associated processes, especially the ETL processes that are used to load it.

Picking Your Metadata

Despite its widespread use, the word *metadata* is not precisely defined. There is no industry standard, and the scope of what is needed is sometimes perceived differently by various people, depending on their requirements. However, there is some common ground. Industry analysts and IT professionals agree that metadata should provide at least some information about the structure of the databases, element names, and the relationship between the different components, including cross-references and impact analysis.

Who Benefits from Metadata?

The benefits of metadata are huge for all those involved in the data warehouse project.

For business users, absent or poor-quality metadata means that they have to figure out the data warehouse's structure and the rules used to load it. Without metadata, this learning curve is steep, and the usage of the data warehouse is restricted. Metadata assists users in understanding the origin of the data: which source systems it came from and which transformations were applied before it was made available in the data warehouse. In this sense, metadata is an essential element of the entire flow, and constitutes an interface between IT developers and business users. Business users are able to drill through metadata, find which rules were developed against production data, and decide to reuse them as templates, instead of trying to create new ones.

For IT professionals, metadata is the basis for efficient change management. The number one benefit is reaped during the maintenance of the data warehouse. Metadata helps to evaluate the impact of changes to the data structures and the loading processes. Indeed, cross-references help the IT team find out which components are accessing a set of data and then analyze the impact of any structural change to existing data transformation processes. Notifying all users of a change that can impact their processes—and which part of these processes is affected—is also fundamental to avoiding execution problems. Metadata dramatically reduces the maintenance workload of the IT team and assists with a quick and smooth evolution of the data warehouse.

The benefits of metadata are huge for all those involved in the data warehouse project.

And for all parties involved, proper metadata management is the key to compliance audits and makes it possible to understand data lineage, data flows, how fields are calculated, and to track the proper execution of ETL processes.

Finding the Right Solution

Finding the right solution to implement metadata may be a challenge. Most ETL tools on the market include some level of metadata management. The key issue here is for this metadata management to be at the core of the ETL design and execution—as opposed to being a way to consolidate metadata afterwards. The ETL tool should most notably be able to:

- Examine the structures and access methods of databases and applications of the information system, and store all the descriptions of these structures inside the metadata repository.
- Allow the definition of data mappings and transformations at an abstract level—using business rules, for example—so that these definitions are de-correlated from the physical implementation of data.
- Provide visualization of all metadata through a user-friendly interface.
- Provide easy navigation among the metadata: dependencies, lineage, data flow maps, and cross-references.

With its business-rules-driven approach to ETL and integration design, along with its centralized metadata repository accessible through a Web-based interface, Sunopsis Data Conductor leverages your metadata and greatly improves the productivity and quality of data warehousing. ■

How To Manage Data Integration Across Your Enterprise

The Three Issues Your Organization Can't Afford to Overlook

By Judy Ko, Director, Enterprise Marketing, Informatica Corporation

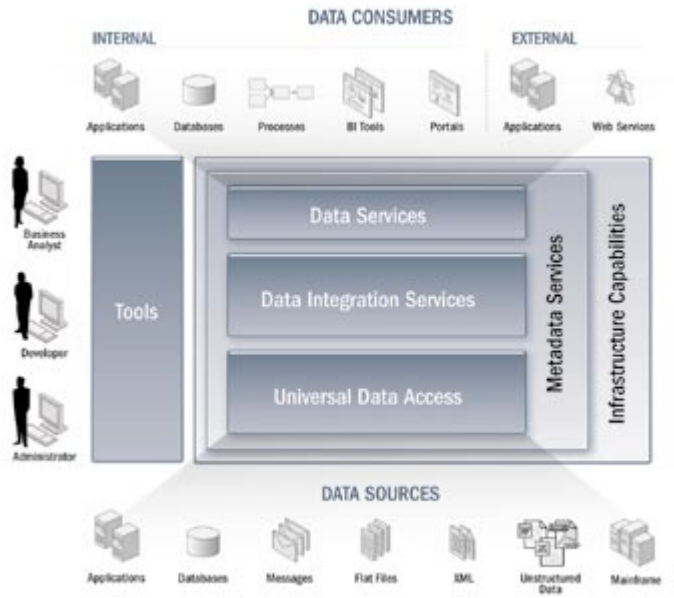
Strategic business initiatives often trigger one or more major IT projects—implementing a single view of the customer, synchronizing multiple operational systems to support an end-to-end business process, or consolidating multiple applications to reduce costs. Yet IT organizations often struggle to deliver on business requirements. What's causing the problem? Data fragmentation.

Data resides in disparate silos throughout the enterprise. The content, quality, structure, and definitions of the data in these silos are as variable as the silos themselves. To ensure that business decisions and operations are based on trustworthy, timely, holistic information, almost every business initiative calls upon IT to access, integrate, and deliver data to the applications and users who need it. In today's stringent regulatory environment, data also needs to be governed properly to meet the auditing requirements of such regulations as the Sarbanes-Oxley Act.

So what's the solution? Data integration.

Data integration allows organizations to access all their fragmented data, create an accurate and consistent view of their core information assets, and easily leverage these assets across the enterprise to drive business decisions and operations.

Historically, IT organizations have used a variety of data integration approaches—extract-transform-load (ETL) tools, hand-coded scripts in conjunction with enterprise application integration (EAI), or application vendors' tools. Over time, this resulted in a proliferation of one-off data integration technologies and complex, brittle IT infrastructures that are costly to manage. As IT organizations take on new initiatives—such as outsourcing a business function—the complexity only increases. To effectively manage data integration across the enterprise, organizations need to look at the problem holistically, taking the following three issues into account:



Informatica provides a single, unified platform, based on a platform-neutral architecture for data integration across the enterprise.

1. Data and IT architecture

Many IT organizations are adopting service-oriented architecture (SOA), and data integration has an important role to play in that architecture. To increase business agility through loose coupling and reusability of data assets, applications and processes must be able to access business-relevant data—wherever it resides, in whatever form required, whenever it's needed—consistently and accurately.

2. Organizational approach

Integration competency centers (ICCs) have emerged as a best practice for enterprise data integration. ICCs are an organizational approach designed to increase agility and reduce implementation costs by promoting reuse, sharing best practices and resources, and establishing common processes and standards for integration. ICCs facilitate cross-enterprise collaboration and coordination for global IT teams, including both internal and external resources such as systems integrators and outsourcers.

3. Data integration technology

To support multiple projects with consistency and maximum reuse, to interoperate within dynamic IT environments, and to ensure robust data governance, organizations need a single, unified enterprise data integration platform that offers:

- Broad access to all enterprise data, regardless of type, structure, or source.
- An open, platform-neutral architecture designed for dynamic, heterogeneous IT environments.
- A single, unified architecture to simplify and accelerate development, deployment, and maintenance.
- Enterprise-class security, scalability, reliability, and availability.
- A shared services approach based on metadata and open standards for transparency, interoperability, and flexibility. ■

Take Your ETL Processes to the Next Level Utilizing p-ETL

By Craig Abramson, Senior Technical Analyst, Syncsort Incorporated

There are numerous ETL (extract-transform-load) products in the marketplace to choose from to help with necessary data processing. But because of the wide range of choices, selecting the right one can be difficult. Many of the products available may have all the bells and whistles, but with the enormous volumes of data that a company may accumulate on a daily basis, this just isn't enough. Companies must consider *performance* when processing data, because slow-running applications waste not only time but money as well. That's why a performance ETL (p-ETL) product is critical.

Companies must consider performance when processing data, because slow-running applications waste not only time but money as well.

There are numerous features a p-ETL product should have in order to complete a specific application, as well as be useful in future applications. Features that tend to be important for most projects include:

- **Performance**—Processing time is critical, so it's important to choose a product that not only has all of the key features needed, but also can complete the data processing as fast as possible.
- **Scalability**—Benchmark tests as well as independent reviews should indicate that a product can easily handle the gigabytes of data generated from large-transaction processing applications.
- **Support**—Since the platform and data formats may change over time, a product should be able to support a variety of data and file formats.
- **Ease of use**—Some products may require complex commands, while others will perform tasks with the click of a mouse. This is especially important when there are a large number of users. A product that's easier to use requires less time to be spent on training or assistance.
- **Price**—Products can range in price from a thousand dollars to over a million. Of course, the higher-end products are not built specifically as ETL products. Also, some products include technical support in the price, while others charge additional fees.
- **Recommendations**—Ask other people in the industry for recommendations on products that they've used, then request free trials of those products.

Once the right p-ETL tool is chosen, it will be used to convert, cleanse, format, and aggregate the data, then pass the data from the warehouse to the appropriate data marts. The data is transformed into a specific format that will help speed querying and reporting time. It can then be accessed and analyzed quickly by the decision makers in the various departments of a company. For example, a marketing analyst can run a report using the sales data mart to determine if there was an increase in product sales soon after the launch of a new marketing campaign. And the easier it is for the decision makers to gather this type of information, the faster they'll be able to turn it into a business advantage.

Beginning the p-ETL Process

In order to take advantage of a p-ETL product, you should begin by evaluating your company's data requirements. This includes deciding what data analysis applications are needed, as well as examining the type of data that is available. You'll also have to identify where each kind of data comes from, how often it is updated, how it is currently being used, and where it can be stored within the company. Then you'll have to decide how you are to going to use it and what must be done to clean the data and transform it. Once this is completed, the next step is to pre-process the data prior to loading it into the data warehouse and database.

Pre-processing allows you to utilize the p-ETL product to reformat the data for seamless integration, and also results in faster and more efficient database and data warehouse loads. Here is a general rule to keep in mind whenever you have a sizable amount of data to load:

> *The format and sequence of load data should be as close as possible to its format and sequence in the database.*

The reason for this rule is simple. Although some database engines load quite efficiently, databases are optimized for query processing and related tasks, not loading. By pre-processing your data

before loading, you leave your database free to do the work it's been designed for, and you gain several performance advantages:

- Your database will load faster
- Indexing will be more efficient
- Data will be ready to use sooner
- Your database will be free to do other work

Transforming data makes it easier and more efficient to access just the information a company needs directly, minimizing the elapsed time of the applications. There are three different levels at which the data transformation processes can take place:

- **Source-level operations**—At this level, database tables can be converted to flat files and vice versa. This allows all of the data to be converted into the same format to speed processing.
- **Record-level operations**—This encompasses such processes as joins, sorts, merges, or just copying records to the appropriate target(s). Before outputting the records, they can be filtered or reformatted for faster access. Aggregations can also be performed at this level.
- **Field-level operations**—This includes data type and format conversions, arithmetic operations, string operations, date-time operations, pattern matching, and conditional operations. These processes help remove extraneous information, allowing only the necessary data to move through the data warehouse.

These kinds of transformations allow data to be cleansed and business rules for data quality to be created. The streamlined data can then be loaded into a data warehouse or other data-intensive application.

Improve Query Response Time

When it comes to data warehouse query performance, speed is everything. Faster performance means needed data and analysis is available to act upon—quickly. In order to achieve this goal, data warehouse experts agree that data aggregates can be a powerful tool. According to Ralph Kimball, author of *The Data Warehouse Lifecycle Toolkit*, data warehouse administrators can "expect anywhere from a tenfold to a thousandfold improvement in runtime performance by having the right aggregates available."

Aggregates are pre-stored summary records held in a data warehouse, and are usually built from the base layer upward, culminating in high-level totals. For example, a data warehouse administrator may decide to build aggregates for sales based on product, store, and time-related totals such as month and year. Tables of aggregates eliminate the time and resources required to summarize data each time it's needed to respond to a query, allowing the query to be answered very efficiently in just a few seconds. A p-ETL product will provide you with the ability to create these aggregates quickly and dramatically improve response time.

When it comes to data warehouse query performance, speed is everything. Faster performance means needed data and analysis is available to act upon—quickly.

Conclusion

An ETL product just isn't enough if it can't complete the necessary processing within the shortest amount of time. Utilizing a performance ETL product can mean the difference between a company that realizes its full potential and one that just keeps missing the mark. After all, p-ETL offers faster availability of data, improved manageability, better decision-making capabilities, and much more. ■

Petabyte Data Warehouses

By Bill Nee, Senior Director, Oracle Corporation

The world's awash in information. It's been estimated that the amount of global data is increasing by about 30 percent per year, with 2006 expected to see a staggering 11 exabytes of total information produced.

Organized information held in individual databases is growing just as fast, and is closing in on the previously mythical one-petabyte (1,125,899,906,842,624 bytes) threshold.

The good news is that vendor databases have grown steadily in sophistication as they've had to respond to the challenges of supporting multi-terabyte databases. But even as technology evolves to support petabyte DWs, IT managers will have some special considerations as they meet this challenge.

Foremost among these is scalability. IT managers must give careful thought to the database being deployed, since even metadata files will be huge. Can the vendor provide proof that it can support very large databases? Do they offer tools that will make management easier? Are data mining and business intelligence tools available that will allow users to easily analyze massive amounts of information? Does the vendor offer parallelism features that will improve performance?

Deploying large databases on SMP (symmetric multiprocessing) or proprietary MPP (massively parallel processing) machines has been, and will continue to be, a popular way of supporting large databases. But the relatively recent option of deploying clusters of commodity servers offers another alternative. A cluster comprises multiple interconnected servers that appear to users and applications as one. The combined processing power of the multiple servers provides greater throughput and scalability than would be available from a single server.

Clusters are inherently high-availability systems, since they provide server redundancy by definition. These systems are also typically less expensive to own and operate, since they consist of an array of cheaper commodity servers. And they are effective: companies have deployed clusters of 64 and even 128 CPUs that have proved capable of handling any current commercial data warehousing workload.

Commercial petabyte databases are now on the horizon, and they'll certainly offer their own unique challenges to customers. But the great thing about our industry has been its ability to meet technical challenges through innovation.

Even with clusters, the cost of amassing the storage and hardware necessary to support a petabyte database will remain high. Though the price of disk storage is decreasing, it's still advisable to use different storage media based on usage. High-speed disk could be used for data that is required on a regular basis; slower, less expensive storage for data that is only occasionally needed; and tape for data that is rarely accessed. Customers should choose a database that can map to both disk and tape and view all media as part of a single database.

High storage costs also underscore the critical need for database partitioning, a powerful feature that allows customers to organize their database into smaller, independent "partitions." Partitioning can keep costs down by allowing a customer to implement a protocol that continually pushes older, unused data onto cheaper storage. It can also ensure data availability and dramatically improve query performance, which are both extremely important when dealing with petabyte-sized data warehouses.

An additional way for customers to lower the cost of their deployment is to utilize open source technology. Unfortunately, open source databases have not yet evolved to the point where they are viable options for most companies. On the other hand, many organizations are using Linux, and a large infrastructure of hardware, software, and services vendors now support it.

The performance of these systems has been very good. As of this writing, a Linux database cluster holds the world record for the TPC-H (data warehousing) 300 gigabyte benchmark, and as one would expect, these systems hold a number of TPC-H price/performance records.

Commercial petabyte databases are now on the horizon, and they'll certainly offer their own unique challenges to customers. But the great thing about our industry has been its ability to meet technical challenges through innovation. With that in mind, it's assured that customers will be able to cross the petabyte boundary just as they did the terabyte mark. ■

The Industry's First BPM System

By John Kopcke, Chief Technology Officer, Hyperion Solutions Corporation

On October 11, 2005, Hyperion introduced System 9, the first business performance management solution to integrate a complete suite of financial management applications with a business intelligence platform into the industry's first BPM system. System 9 is the result of a three-year initiative to dramatically transform and simplify the way users interact with business performance management and BI software and the way performance management solutions are deployed, managed, and maintained across the enterprise.

The seeds of System 9 were present when Hyperion entered the performance management market in the early 2000s, and it is a natural next step in the evolution of the category. Then, fragmented decision support systems on top of a complex, multi-vendor transactional environment populated by multiple data sources were the norm. From both the business user's and the IT department's point of view, it was total anarchy.

The dirty little secret is that in many enterprises, things look pretty much the same today. Vendors are still offering point products, with few tying them together as part of an overall solution.

Customers visiting our briefing center confirm this fact almost every day. They tell us that they are trying to simplify their transactional systems to get fewer moving parts. But they also tell us that their reality is a multi-vendor world, and that they need help pulling it all together.

System 9 is a single system for delivering business performance management across the enterprise. It includes shared business data and processes along with other common services. It creates more interoperability between the applications and activities in the performance management cycle, and enables even greater alignment across the enterprise.

The core of System 9 is financial management applications and our BI platform, which Hyperion has been evolving for some time. What's new about System 9 is that we have integrated them into a single system. We have done this by developing a System 9 Workspace that unifies applications, data, and reports into a single, easy to learn and use environment for business users. And we have developed Foundation Services that make deploying, managing, and maintaining performance management solutions easier and less costly for IT.

> Fragmented decision support systems on top of a complex, multi-vendor transactional environment populated by multiple data sources were the norm … it was total anarchy.

The focal point of the innovation behind System 9 is the Workspace. With it, users can address all of their performance management and query and reporting needs within a single thin-client environment that is personalized and optimized for their unique needs. As a result, users are no longer forced to shuffle between applications to get answers to their business questions or to take action. In addition, the System 9 Workspace reduces the amount of training that is required for business users to learn business performance management and BI software. This translates into a reduction in the time IT must spend supporting users.

Hyperion System 9 is a single, comprehensive, yet flexible system that has many advantages:

- **Increases user productivity.** Through the System 9 Workspace, business users can access and use all elements of their business performance management solution—applications and BI tools—based on their role in the enterprise.
- **Reduces business risk.** Advanced data integration and management services in System 9 enable enterprises to more easily and cost-effectively create one version of the truth.
- **Lowers total cost of ownership.** Integrating applications and a BI platform means System 9 retains modularity and flexibility, easing management.

These are compelling benefits that are available to all Hyperion customers, regardless of which of our products they currently use.

Our commitment to evolving the BPM System will not end with System 9. Already the most comprehensive business performance management solution ever and the first to integrate financial management applications with BI into a single system, System 9 is just the beginning. ■

The "Right Time" for Operational Business Intelligence?

By Lisa Dreyer, Director of Product Marketing, Sybase, Inc.

For many organizations, the business intelligence (BI) system enables power users and skilled business users to report on and analyze business data so that organizations can reduce costs and increase revenues. Many companies and government agencies routinely use a BI system for strategic and tactical decision making. Analysts look at what's happened in the past, figure out why it's happened, and make changes to future strategies.

What's Changed?

An intensely competitive global climate is forcing organizations to react faster to changing business conditions and customer demands. Further, it's requiring that the BI system drive and optimize critical operations—daily or intraday. This smarter use is what's termed *operational business intelligence* or operational BI. The objective of operational BI is to make timelier business decisions; therefore, it has a close relationship to right-time BI.

Right-time BI is all about delivering the right information in the right format to the right people at the right time for decision-making purposes. Right-time BI optimizes ("right-sizes") the time latency between when a business event occurs and when appropriate action is taken.

So, why are organizations now getting serious about achieving this goal? Organizations are under more careful—and public—scrutiny than ever before. The media and the competition are looking at new global compliance regulations as well as highly publicized business results: the wins… and the losses. Customer expectations are high; margins for error are shrinking; and tolerance for poor business intelligence is low.

The goal of operational BI is to close the gap between analytical applications and operational applications, creating a closed-loop process. Ideally, this brings the benefits of business intelligence to a broader population within an organization. People and processes in operational departments can now take advantage of the power of this tool. Business intelligence is no longer just for power users and analysts.

What Are the Requirements?

To achieve the competitive benefits of right-time BI (and ultimately operational BI), organizations need to select underlying technology that will support the basic requirements of the environment: speed, scalability, flexibility, low operating cost, and fast backup and recovery.

- **Speed**—Operational BI systems must provide answers to ad hoc queries within "right-time" requirements. Applications may be on the Web, and are often customer-facing. Tolerance for analysis latency is measured in seconds and minutes, with no failed queries. Ultimately, to stay ahead of the game, companies need faster answers internally, in addition to faster customer services than their competition.

- **Scalability**—Operational BI must be highly available to everyone involved in the key operations of daily business. The system must be able to incrementally scale to thousands of concurrent users and to terabytes of data, while maintaining query speeds of seconds or minutes.

- **Flexibility**—Operational BI systems must support a variety of schemas, configurations, and queries to meet business needs. To optimize ROI, deployment must be fast and must be based on standard hardware and operating systems. For ease of integration, the system must be compatible with existing applications and tools. Remember, maintaining nonstandard platforms adds costs and complexity to the organization.

- **Low operating cost**—Look for an analytics server, like Sybase IQ, that provides the price/performance to achieve your operational BI goals while reducing overall costs. This quotient encompasses reducing the costs of tuning, administration, support, and maintenance; reducing storage costs by compressing data; and using lower-cost platforms like Linux.

- **Fast backup and recovery**—Operational BI systems are mission-critical to organizations. The cost of downtime to the business or government agency can be enormous, especially if the system hosts a customer-facing or Web-based service critical to business operations. Look for a system designed to support an efficient backup and recovery plan.

Conclusion—The "Right Time" for Operational BI Is Now

Strategic and tactical BI is well entrenched in most organizations, and the biggest growth area in BI over the next few years will be in operational BI. This type of BI promises significant business benefits and extends to a wide user audience. Make sure you are making technology decisions today that will support the needs of this next generation of BI. ■

Taking Data Quality to the Enterprise through Data Governance

By Philip Russom, Senior Manager of Research and Services, TDWI

The Scope of Data-Quality Initiatives

Historically applied to isolated silos in departments or single databases, data quality is progressively applied with more breadth across enterprises. TDWI research shows that a surprisingly large number of organizations (39%) are already applying data quality in some form to the "whole enterprise" (see Figure 1). An additional 26% apply it in a "single department that spans the enterprise," such as the IT and marketing departments. Meanwhile, low percentages continue the older tradition of applying data quality mostly in "a single department" (14%) or "a single business unit" (13%).

The gist of the market data is that many organizations are well down the road to enterprise-scope use of data-quality techniques and practices—enterprise data quality (EDQ). Users interviewed for this research reported similar progress, corroborating the survey data. But most interviewees quickly added that they had only recently arrived at EDQ, typically in 2003 or 2004. Hence, the data-quality marketplace and user community has only recently crossed the line into EDQ. TDWI suspects that many more organizations are on the cusp, and will cross into EDQ in coming years. Of course, occasional departmental usage will continue alongside EDQ.

One of the reasons data-quality usage is spreading is that it is often piggybacked atop related initiatives that carry it across the enterprise. Just about any data-intense initiative or software solution will ferret out data-quality problems and opportunities. IT and business sponsors have realized this over time, so it's become commonplace to include a data-quality component in initiatives for governance (49%), CRM (42%), marketing campaigns (34%), compliance exercises (35%), and supply chain management (16%) (see Figure 2).

Anecdotal evidence suggests that data quality gives these initiatives better planning, a more predictable schedule, and a higher-quality deliverable. Likewise, data-quality software is often integrated with software for other solutions, like data warehousing and business intelligence (79%), customer data integration (38%), migrations and consolidations (35%), and master data management (35%) (see Figure 3).

Most Data-Quality Trends Lead to Enterprise Use

The practice of data quality is in a state of transition, as its every aspect is currently evolving due to a strong trend. Most of these trends have a general effect

What is the scope of your data-quality initiative?

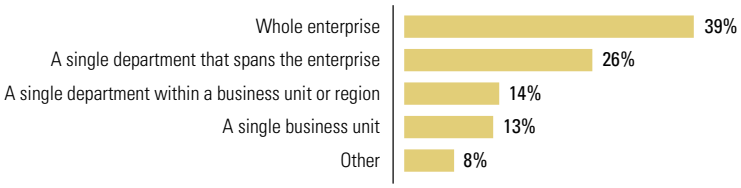


Figure 1. *Based on 569 respondents (only those who have a data-quality initiative).*

Which business initiatives does your data-quality initiative support?

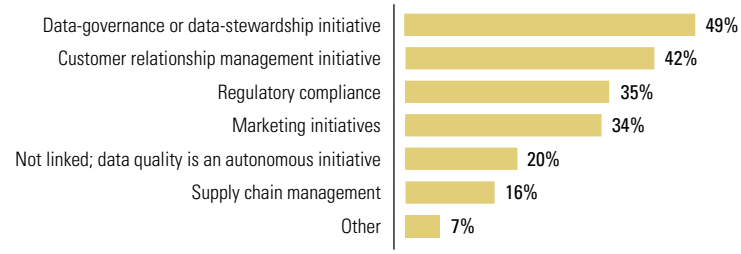


Figure 2. *Based on 1153 responses from 569 respondents.*

Which software solutions does your data-quality initiative support?

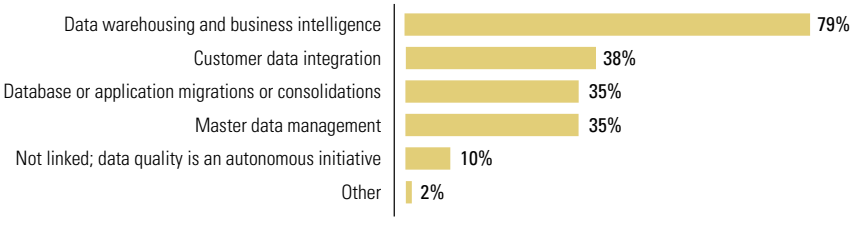


Figure 3. *Based on 1137 responses from 569 respondents.*

in common—they result from or cause a broadened use of data-quality tools and practices across an enterprise (see Figure 4).

From departmental solutions to enterprise initiatives. Most IT directors interviewed for this report spoke of their first data-quality projects as supporting data warehousing or marketing functions like direct mail. A few mentioned product data issues, like catalog matching and cleansing. Regardless of the isolated silos where they started, data-management professionals now pursue the quality of data broadly across their organizations. But many are at a fork in the road: either they keep deploying data-quality silos in more departments, or they fall back and regroup into a centralized team that gains efficiency and consistency across all efforts. TDWI recommends the second route, because it establishes a structure that leads to further progress in the long run.

From technical users to business users. Long story short, the data-quality user community gets more diverse all the time. At one extreme, technical users design large, scalable solutions and programs for matching and consolidation rules. At the other extreme, semi-technical marketers handle name-and-address cleansing and other customer data issues. With the rise of stewardship years ago and data governance recently, there's a need to support business users who are process and domain experts, with little or no technical background. This trend affects the tools that vendors provide; most were designed for one of these user constituencies, and now must support them all. Likewise, organizations must staff their data-quality initiatives carefully to address all these user types, their needs, and their unique contributions.

From point products to tool suites. TDWI defines data quality as a collection of many practices, which explains why few vendors offer a single "data-quality tool." Instead, most offer multiple tools, each automating a specific data-quality task like list scrubbing, fuzzy matching, geo-coding, and so on. As organizations broaden data-quality usage, they use more of these point products, then suffer the lack of interoperability, collaboration, and reuse among them. To supply this demand, vendors have worked hard to integrate their point products into cohesive suites.

Data-quality products and practices are evolving quickly…

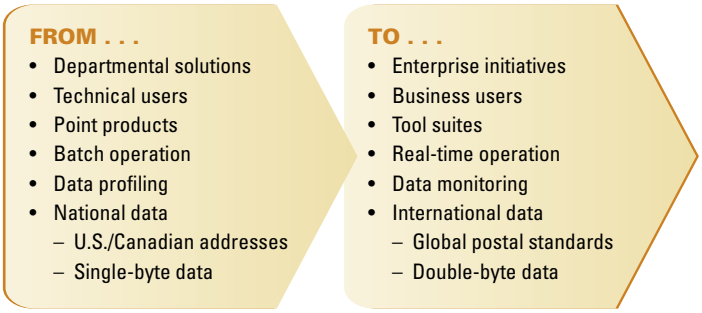


Figure 4. *The evolution of data-quality products and practices.*

From batch to real-time operation. Some data-quality tasks are still best done in batch, like list scrubbing and matching records in large datasets. However, given that data entry is the leading source of garbage data, there's a real need for real-time verification, cleansing, and enhancement of data before it enters an application database. As with most data-management practices, data-quality software has evolved to support various speeds of "right-time" processing.

From data profiling to data monitoring. These are similar in that each results in an assessment of the quality of data. On the one hand, data profiling has the additional step of data discovery and is done deepest prior to designing a data-quality solution. Monitoring, on the other hand, is about measuring the quality of data frequently while a data-quality solution progresses, so stewards and others can make tactical adjustments to keep a quality initiative on plan. The trend is to do profiling more deeply, then embrace monitoring eventually. TDWI strongly recommends both.

From national to international data. Companies that are multi-national or have a multi-national customer base have special problems when expanding data-quality efforts across a global enterprise. Name-and-address cleansing is a straightforward task when done for U.S. and Canadian addresses and postal standards; yet it becomes quite complex as you add more languages, national postal standards, and information structures (like Unicode pages and double-byte data). Users must deal with these and other issues as they take data quality to an international enterprise.

Data Governance and Enterprise Data Quality

TDWI data shows that many organizations are practicing enterprise data quality in some sense. The catch, however, is that practices from isolated areas (like data warehousing or marketing campaigns) aren't automatically successful on an enterprise scale. Accomplishing anything at the enterprise level requires close cooperation among IT and business professionals who understand the data and its business purpose and have a mandate for change. To achieve this, an organization can establish a data-governance committee according to the following definition:

> *When an organization views data as an enterprise asset (transcending the data warehouse and spanning the whole organization), it establishes an executive-level* ***data-governance*** *committee that oversees data stewardship across the organization. Depending on the scope of a data-governance initiative, it may guide related initiatives, like data quality, data architecture, data integration, data warehousing,*

metadata management, master data management, and so on.

Distinctions between stewardship and governance are thin in some cases. But TDWI sees data stewardship as a local task that protects and nourishes specific data collections for specific purposes (like a data warehouse for business intelligence or marketing databases for direct mail). Data governance is a larger undertaking that exerts control over multiple business initiatives and technology implementations, to unify these through consistent data definitions and gain greater reuse for IT projects and business efforts. The two can work together, in that a data-governance committee can be a management level that coordinates multiple data-stewardship teams. In a few companies, data governance is a subset of an even larger corporate governance initiative.

The most critical success factor with governance is *mandate*. Governance bodies and stewards must exert change on business and technical people—who own the data and its processes—when opportunities for improvement arise. The most effective mandates come from a high-level executive. Without a strong mandate for change and an attentive executive sponsor, stewardship and governance deteriorate into academic data profiling exercises with little or no practical application.

The State of Data-Governance Initiatives

TDWI's data-quality surveys asked: "Who is responsible for data quality in your organization?" In both 2001 and 2005, the data warehouse team and IT bubbled to the top of the list (47% and 43% in Figure 5). This makes sense from a technology viewpoint, in that these are the technical people long involved in data quality. But it gives technology priority over business, whereas the two must collaborate in a stewardship or governance program. Respondents ranked business analysts and power users in third place (30%), followed by the "cross-functional team from business and IT" (28%), a description that includes both sides, as in our definition of data governance. So, respondents recognize that responsibility for the quality of data must be shared by some kind of cross-functional team. But the fact that "data-quality analysts" and "data stewards" ranked even lower than "front-line workers" (whose data entry is the leading cause of garbage data), indicates that sharing responsibility through governance and stewardship is still rare.

When asked about data-governance initiatives, a disappointing 8% reported having deployed one, while 42% have "no plans" (see Figure 6). TDWI then asked respondents (except those with "no plans") to rank the effectiveness of the steering committees, degree of executive involvement, and usefulness of policies and processes in their data-governance initiative. The majority ranked all three areas as "moderate," meaning there's plenty of room for improvement.

These disappointing responses are most likely due to data governance being a relatively new approach, coupled with the fact that many organizations seem to be on the cusp—they've stretched data-quality practices over the enterprise in a disconnected way and now it's time to control them to ensure consistency and efficiency, whether the control is via stewardship, governance, or centralized IT services. Various forms of data governance will, no doubt, disseminate as more organizations come off the cusp.

Anatomy of Data-Governance and Stewardship Programs

Staffing and management hierarchy for data-governance and stewardship programs will vary according to each organization's unique structure and

Who is responsible for data quality in your organization?

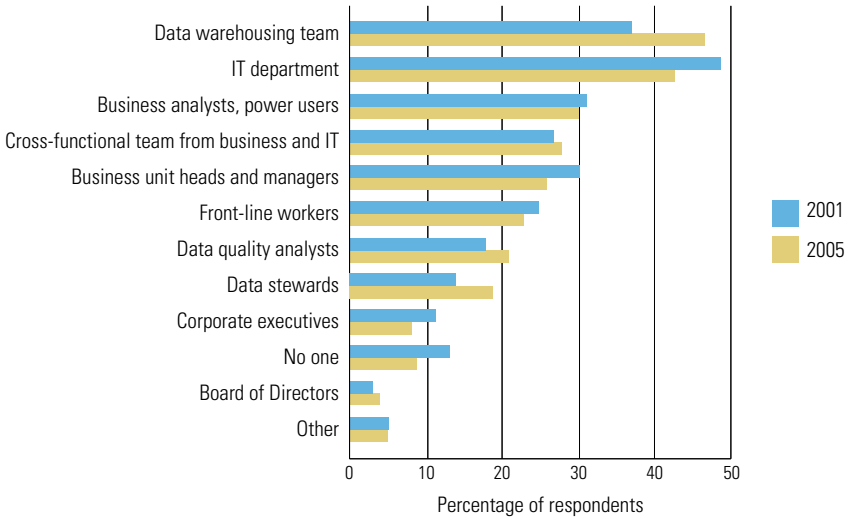


Figure 5. *Based on 1701 responses from 647 respondents in 2001, 1957 responses from 750 respondents in 2005.*

What's the status of your organization's data-governance initiative?

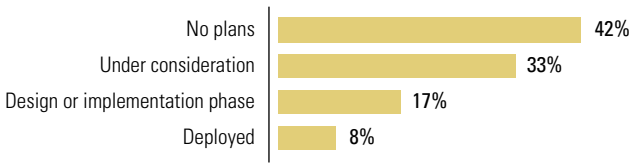


Figure 6. *Based on 750 respondents in 2005.*

needs. The following description—a composite drawn from multiple interviewees—illustrates the requisite parts:

- **A domain steward is assigned per business unit.** Domain stewards work directly with the line-of-business manager who owns the data and the IT manager who administers it. Each steward has a mandate to change the process and structure of any business, person, or IT system, if that's what it takes to improve data. Note that most changes proposed should be business oriented, with business value as a goal of any data-quality work that gets done. Without demonstrable value, it's unlikely the work will get approved or done.

- **A corporate steward manages a group of domain stewards.** This management hierarchy helps related domain stewards collaborate. And the corporate steward provides domain stewards with additional clout to help domain stewards enforce their mandates.

- **A governance committee consists of miscellaneous managers**. These include corporate stewards, corporate sponsors (CxOs and SVPs), and miscellaneous IT and line-of-business managers, as needed. This committee sets top-down strategic goals, coordinates efforts, and provides common definitions, rules and standards, which apply to data structures, access, and use across the entire enterprise.

Figure 7 shows how the layers of stewardship and management may roll up into a data-governance committee. Dark arrows represent direct reports, while gray arrows represent significant interactions outside the reporting structure of the organization.

Conclusions and Recommendations

Data quality proved itself in its data warehouse and direct mail origins, and has now moved beyond these into enterprise data quality, where it is applied in many departments for many purposes. In fact, most trends result from or cause a broadened use of data quality tools and practices across the enterprise. Plus, data quality is now *de rigueur* as a component of various business initiatives and software solutions. While this broadening is good for the data, it's challenging for the organization, which must adjust its business processes and IT org chart to adapt.

Get ready for enterprise data quality. It will improve many business and technical processes, if you're open to its diversity and give it necessary organizational structure.

- **Embrace the diversity of data-quality practices.** Many organizations need to move beyond name-and-address cleansing, data warehouse enhancement, and product catalog record matching. These are useful applications, but are narrow in scope. The lessons learned and skills developed for these can be leveraged in other data-quality applications across the enterprise.

- **Address enterprise data quality.** The data-quality initiatives of 39% of survey respondents already address the "whole enterprise." Follow their lead into enterprise usage, but resist the urge to deploy data quality in isolated pockets of software tools and IT personnel. Some kind of centralization can improve personnel allocation, project reuse, and data consistency.

- **Give EDQ required organizational structure through data governance.** EDQ's chances of large-scale, long-term success are limited without a support organization, whether its form is a data governance committee, a data stewardship program, or a data quality center of excellence. Another key requirement is a strong mandate supported by a prominent executive sponsor. ■

Philip Russom is TDWI's senior manager of research and services. Before joining TDWI in 2005, Russom was an industry analyst covering data warehousing, business intelligence, and integration technologies at Forrester Research, Giga Information Group, and Hurwitz Group. He also ran his own business as an independent industry analyst and consultant, and was contributing editor with Intelligent Enterprise *and* DM Review *magazines. You can reach him at prussom@tdwi.org.*

Excerpted from the full April 2006 report. TDWI appreciates the sponsorship of Business Objects, Collaborative Consulting, DataFlux, DataLever, Firstlogic, IBM, Informatica, Similarity Systems, and Trillium Software.

To download the full report, visit www.tdwi.org/research/reportseries.

Possible Organizational and Report Structure for Data Governance

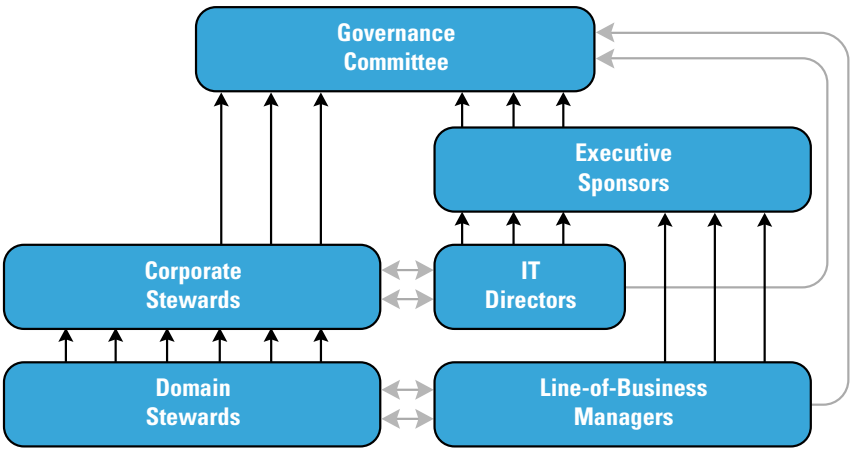


Figure 7. *Stewardship and management roll up into data governance.*

About TDWI

Mission

TDWI™, a division of 101communications, is the premier provider of in-depth, high-quality education and research in the business intelligence (BI) and data warehousing (DW) industry. Founded in 1995, TDWI is dedicated to educating business and information technology professionals about the strategies, techniques, and tools required to design, execute, and maintain successful BI and DW projects. Within the community it serves, it provides a comprehensive resource for professional development and fosters knowledge sharing and the advancement of research. TDWI sponsors and promotes a worldwide Membership program; quarterly educational conferences; regional educational seminars; onsite courses; certification; solution provider partnerships; an awards program for best practices in DW, BI, and other innovative technologies; resourceful publications; an in-depth research program; and a comprehensive Web site (www.tdwi.org).

Membership

As the BI and DW field continues to evolve and develop, it is necessary for information technology professionals to connect and interact with one another. TDWI provides opportunities for such individuals to learn from each other, network, share ideas, and respond as a collective whole to the challenges and opportunities in the industry.

TDWI's Membership includes 5,000 BI, DW, and information technology (IT) professionals from *Fortune* 1000 corporations, consulting organizations, and governments in 45 countries.

TDWI Member Benefits

- Quarterly *Business Intelligence Journal*
- Annual *TDWI Salary, Roles, and Responsibilities* Report
- Annual *BI Benchmark Report*
- Quarterly *Ten Mistakes to Avoid* series
- Biweekly *TDWI FlashPoint* e-newsletter
- Quarterly Best Practices Research Reports
- Quarterly Technology Market Reports
- Quarterly TDWI Member Newsletter
- Consulting calls (Team Memberships only)
- MemberConnect, an online networking forum
- Access to password-protected, Members-only online content at www.tdwi.org, including TDWI research, publications, and archives
- Special discounts on event registration, certification exams, books, and TDWI merchandise
- Discounts on custom research (Team Memberships only)

Membership in TDWI is available to all BI, DW, and IT professionals for an annual fee of $275 ($325 outside the U.S.). TDWI also offers Team Membership with special benefits for organizations that register multiple individuals as TDWI Members.

To learn more about TDWI Membership, visit www.tdwi.org/Membership.

TDWI World Conferences

TDWI's quarterly educational conferences offer a rigorous training environment in a conference setting and incorporate an intensive, full-day course format and extended course programs that span several days. All courses are selected for content and quality and are evaluated by attendees to ensure adherence to TDWI standards. TDWI attendees gain knowledge through a full-day format that prepares them to quickly apply their new knowledge toward a more successful BI or DW initiative. TDWI's unique conference model has set a standard for education that is now imitated at other conferences.

Seminar Series

The TDWI Seminar Series makes it easier for busy professionals and project teams to receive training closer to home and in less time. These courses assist managers in bringing new team members up to speed quickly and cost effectively.

Onsite Courses

For the ultimate in convenience, TDWI can bring top-rated certificate courses to your location. The benefit of onsite education is that the course can be customized to meet a group's particular challenges and issues. It also provides a quick and easy way to give all team members an equivalent understanding of the core issues and concepts of BI and DW and how they apply to a company's environment.

Certification

Certified Business Intelligence Professional (CBIP), a true test-based certification program, is offered in five key specialties for BI success: Leadership & Management, Business Analytics, Data Analysis & Design, Data Integration, and Administration & Technology. Learn more at www.tdwi.org/Certification/CBIP.

About 101communications

Founded in 1998, 101communications is an integrated media company in the business-to-business market aimed at the many specialized targets within the greater information technology community. 101communications' portfolio includes 11 publications, 23 Web sites, 36 e-newsletters, and 44 conferences, trade shows, and seminars worldwide in eight areas: business intelligence, education, government, network and enterprise, office technology recycling, software solutions, certification, and compliance. ■

Actuate Corporation
701 Gateway Boulevard
South San Francisco, CA 94080
650.837.2000
Fax: 650.827.1560
info@actuate.com
www.actuate.com

Actuate Corporation is the world leader in enterprise reporting and performance management applications that empower 100 percent of users to achieve breakthrough corporate performance. Actuate provides the most scalable, reliable, flexible, and high-performing reporting capabilities for every application in the enterprise. Customers use Actuate to deliver information in context to users inside and outside the firewall as performance management and customer self-service applications, managed spreadsheet applications, and Java reporting applications. Actuate has over 3,500 customers globally in a range of industries, including banking, insurance, manufacturing, communications, and government.

Business Objects
North American
Corporate Headquarters
3030 Orchard Parkway
San Jose, CA 95134
800.527.0580, 408.953.6000
Fax: 408.953.6001
www.businessobjects.com/
company/contact_us/form.asp
www.businessobjects.com

Business Objects is the world's leading BI software company, helping organizations gain better insight into their business, improve decision making, and optimize enterprise performance. The company's business intelligence (BI) platform, BusinessObjects™ XI, offers the industry's most advanced and complete platform for reporting, query and analysis, performance management, and information management including data integration. BusinessObjects XI includes Crystal Reports®, the industry standard for enterprise reporting. Business Objects also has the industry's strongest and most diverse partner community, with more than 3,000 partners worldwide. In addition, the company offers consulting and education services to help customers effectively deploy their business intelligence projects.

Collaborative Consulting
Ten Tower Office Park, Ste. 218
Woburn, MA 01801
877.376.9900
Fax: 781.376.9800
info@collaborative.ws
www.collaborative.ws

Collaborative Consulting is a leading professional services organization that specializes in optimizing its clients' business and technology capabilities. We combine exceptional business knowledge and market-leading technology expertise with an effective partnership approach, allowing us to understand and solve even the most complex business problems. And, by aligning business and technology initiatives, we help clients achieve superior, cost-effective business solutions. Founded in 1999, Collaborative provides operational consulting, program management, data services, and technology services for clients across the U.S., with headquarters in Woburn, MA. Collaborative's Web site is www.collaborative.ws.

DataFlux Corporation
940 Cary Parkway, Suite 201
Cary, NC 27513
877.846.3589
www.dataflux.com
sales@dataflux.com

DataFlux Corporation, a wholly owned subsidiary of SAS, enables companies to analyze, improve, and control their data through an integrated technology platform. Through its enterprise data quality integration solutions, companies can build a solid information foundation that delivers a unified view of customer, product, or supplier data. DataFlux helps customers enhance the effectiveness of their data-driven initiatives, including customer data integration (CDI), enterprise resource planning (ERP), product data management (PDM), and compliance.

DataFlux provides a full range of solutions and services and has more than 750 customer installations worldwide. To learn more about DataFlux, visit www.dataflux.com.

DATAllegro, Inc.
65 Enterprise
Aliso Viejo, CA 92656
949.330.7690
info@DATAllegro.com
www.DATAllegro.com

DATAllegro's data warehouse appliances lead the industry in price and performance. DATAllegro's patent-pending technology enables significantly faster and more complex queries than any other data warehouse appliance on the market. DATAllegro P3™ is a high-performance data warehouse handling up to three terabytes of user data, and is priced at $150,000 per terabyte. DATAllegro C25™ is a high-capacity data warehouse holding up to 25 terabytes of user data, and is priced at $20,000 per terabyte. With its breakthrough technology, DATAllegro has set a new standard in data warehouse price/performance. For more information on DATAllegro, go to www.DATAllegro.com.

Hyperion Solutions Corporation
5450 Great America Parkway
Santa Clara, CA 95054
408.744.9500
Fax: 408.588.8500
www.hyperion.com

Hyperion Solutions Corporation is the global leader in business performance management software. More than 11,000 customers rely on Hyperion software to provide visibility into how their businesses are performing and to help them plan and model to improve that performance. Using Hyperion software, customers collect data, organize and analyze it, then communicate it across the enterprise. Hyperion offers the industry's only business performance management solution that integrates financial management applications with a business intelligence platform into a single system.

Named one of the *Fortune* 100 Best Companies to Work For (2004), Hyperion serves global customers in 45 countries. A network of more than 600 partners provides the company's innovative and specialized solutions and services. Hyperion generated revenues of $703 million for the fiscal year that ended June 30, 2005, and is traded under the Nasdaq symbol HYSL. For more information, please visit www.hyperion.com.

HyperRoll Inc.
1300 Crittenden Lane, Ste. 104
Mountain View, CA 94043
650.962.9192
Fax: 650.962.2501
www.hyperroll.com

HyperRoll Inc. provides software that enables business intelligence solutions to become massively scalable. HyperRoll delivers previously unattainable levels of speed, data volume, and business dimensions. By using HyperRoll in conjunction with existing infrastructure, HyperRoll provides customers with lightning-fast reporting capabilities and allows customers to perform more sophisticated queries. HyperRoll combines patented aggregation, caching, and compression algorithms to deliver immediate and substantial improvements in business intelligence performance.

IBM Corporation
1133 Westchester Avenue
White Plains, NY 10604
800.IBM.4YOU
www.ibm.com/contact/us/
www.ibm.com/bi

IBM is the world's leading IT provider. For nearly a century, IBM has been enabling companies to make better-informed decisions by helping businesses to generate, manage, and extend their enterprise data warehouse. IBM combines industry-leading eServers with the most extensive portfolio of information management software in the industry, from world-class database software to information integration, master data management, analytics, Web services, and more, to deliver solutions ranging from department-level datamarts to massive enterprise data warehouses.

Informatica Corporation
100 Cardinal Way
Redwood City, CA 94063
800.653.3871, 650.385.5000
Fax: 650.385.5500
myinformatica@informatica.com
www.informatica.com

Informatica Corporation delivers data integration software and services to solve a problem facing most large organizations: the fragmentation of data across disparate systems. Informatica helps organizations gain greater business value from their information assets by integrating their enterprise data. Informatica's open, platform-neutral software reduces costs, speeds time to results, and scales to handle data integration projects of any size or complexity. With a proven 13-year track record of success, Informatica helps companies and government organizations of all sizes realize the full business potential of their enterprise data. That's why Informatica is known as the data integration company.

MicroStrategy
1861 International Drive
McLean, VA 22102
888.537.8135
Fax: 703 848 8610
info@microstrategy.com
www.microstrategy.com

MicroStrategy is the global leader in business intelligence (BI) technology. Founded in 1989, MicroStrategy provides integrated reporting, analysis, and monitoring software that helps leading organizations worldwide make better business decisions every day. Companies choose MicroStrategy for its advanced technical capabilities, sophisticated analytics, and superior data and user scalability.

With over 15 years of industry experience, thousands of customer successes, and a reputation for innovation and leadership, MicroStrategy is the safest choice for your business intelligence investment. More information about MicroStrategy is available at www.microstrategy.com.

Oracle Corporation
500 Oracle Parkway
Redwood Shores, CA 94065
800.ORACLE1 or
650.506.7000
www.oracle.com/corporate/contact/

Oracle is the world's leading supplier of software for information management, and the world's second largest independent software company. Oracle software can be found in nearly every industry around the world and in the offices of 98 of the *Fortune* 100 companies. Oracle is the first software company to develop and deploy 100 percent Internet-enabled enterprise software across its entire product line: database, business applications, and application development and decision support tools. Oracle's offerings include new software licenses, software license updates, and product support and services, which include consulting, advanced product services, and education.

Sunopsis
6 Lincoln Knoll Lane, Ste. 100
Burlington, MA 01803
781.238.1770
Fax: 781 238 1771
info-us@sunopsis.com
www.sunopsis.com

Sunopsis is the leading provider of Simply Faster Integration software, optimized for data warehousing and business intelligence projects. Sunopsis products leverage a business-rules approach to data integration and use a unique E-LT (Extract–Load & Transform) approach to their execution. This leverages the existing database engines and executes the ETL processes where they can run the most efficiently, achieving 10 times better performance than traditional ETL tools at a lower cost. Quick to deploy, Sunopsis is the most cost-effective solution on the market today. Find out why leading companies throughout the world have chosen Sunopsis: visit our Web site at www.sunopsis.com/us/ww21.

Sybase, Inc.
One Sybase Drive
Dublin, CA 94568
800.8.SYBASE
infobi@sybase.com
www.sybase.com

Sybase is the largest global enterprise software company exclusively focused on managing and mobilizing information from the data center to the point of action. Our highly optimized analytics engine, Sybase IQ, is designed specifically to deliver dramatically faster results for mission-critical business intelligence, analytics, and reporting solutions on standard hardware and software platforms. Sybase IQ delivers unsurpassed query performance and storage efficiency for structured and unstructured data. Sybase IQ combines extraordinary speed and agility with low total cost of ownership, enabling enterprises to perform analysis and reporting that was previously impossible, impractical, or cost prohibitive.

Syncsort Incorporated
50 Tice Boulevard
Woodcliff Lake, NJ 07677
201.930.9700
Fax: 201.930.8290
info@syncsort.com
www.syncsort.com

Syncsort Incorporated is a leading developer of high-performance backup, data management, and data warehousing software for mainframe, UNIX, and Windows environments. For over 35 years, Syncsort has built a reputation for superior product performance and technical support. An independent market research firm has named Syncsort as one of the top Data Warehouse 100 Vendors for seven years in a row, and the company was recently chosen as the leading provider of data acquisition and integration products. Syncsort's products are used to back up and protect data in distributed environments, speed data warehouse processing, improve database loads, and speed query processing.

TDWI Partner Members

These solution providers have joined TDWI as special Partner Members and share TDWI's strong commitment to quality and content in education and knowledge transfer for business intelligence and data warehousing.

baseline CONSULTING

Hyperion

ORACLE

Business Objects

IBM

sas The Power to Know.

COGNOS THE NEXT LEVEL OF PERFORMANCE

INFORMATICA The Data Integration Company

SILVERTRAIN

CONNECT: The Knowledge Network

KNIGHTSBRIDGE

Sunopsis

Conversion Services International Data Warehousing Specialists.

Microsoft

Teradata a division of NCR

DATAFLUX A SAS COMPANY

MicroStrategy Best In Business Intelligence

HARTE HANKS TRILLIUM SOFTWARE

DecisionPath CONSULTING

NETEZZA The Power to Question Everything

UNISYS Imagine it. Done.

What Works

BEST PRACTICES IN BUSINESS INTELLIGENCE AND DATA WAREHOUSING

Volume 21

5200 Southcenter Boulevard
Suite 250
Seattle, WA 98188

Fax: 206.246.5952
info@tdwi.org
www.tdwi.org

FREE Business Intelligence & Data Warehousing White Papers!

TDWI has gathered a collection of the most outstanding white papers available for the business intelligence and data warehousing industry. Please select as many white papers as you would like to receive in any of the categories below. These will be sent directly to you—with no obligation. (To select your desired white papers, please click on the titles. Note that you must be connected to the Internet.) You can also access these white papers on our Web site at http://www.tdwi.org/publications/whitepapers/ww21whitepapers.aspx

Administration and Operations

1. The Business Case for Database Partitioning

Analytic Apps and Development Tools

2. Defining an Enterprisewide Business Performance Management Architecture

Business Analytics

3. Enterprise Business Intelligence: Improving Corporate Performance Through Integrated Reporting, Analysis, and Monitoring
4. Right-Time Business Intelligence: Optimizing the Business Decision Cycle
5. The Managed Spreadsheet Environment

Data Integration

6. Addressing Data Quality at the Enterprise Level
7. Data Integration: Creating a Trustworthy Data Foundation for Business Intelligence
8. Data Integration in a Service-Oriented Architecture
9. Information Integration: New Capabilities in Data Warehousing for the On Demand Business
10. Is ETL Becoming Obsolete? E-LT is a Better Approach
11. Master Data Management: Creating an Enabling Platform for Business Integration
12. Optimizing ETL Operations: The Key to Timely Business Decisions Based on Database and Data Warehouse Results
13. The Challenges of Customer Data Integration (CDI)

Infrastructure

14. Joins and Aggregates Are Critical for Achieving Faster Data Retrieval in a Dimensional Data Warehouse
15. Performance Optimization: A Necessary Component of EIM
16. The Truth About Data Warehouse Appliances