

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

Ahold and Netezza

Alaska Air Group and Oracle

BP and XLCubed

Cancer Care Ontario
and MicroStrategy

City of Troy and ESRI

DATALlegro

Good Technology and Syncsort

IXIS CIB and Sybase

Mentor Graphics and
Hyperion Solutions Corporation

Nationwide Insurance
and Informatica

RLPTechnologies and DataFlux

Schibsted Search and FAST

UGS PLM Solutions and arcplan

Université du Québec à Montréal and
Business Objects

in sponsorship with TDWI

Volume 22



tdwi
THE DATA WAREHOUSING INSTITUTE

TDWI proudly presents volume 22 of *What Works: Best Practices in Business Intelligence and Data Warehousing*. This compendium of content provides an information resource for better understanding the tools, technologies, and methods that are central to this growing industry.

CASE STUDIES The goal of *What Works* is to provide a snapshot of the most innovative business intelligence and data warehousing implementations in the industry today. The case studies included in this volume demonstrate the power of BI for industries ranging from retail to education, insurance to transportation.

LESSONS FROM THE EXPERTS *What Works* also 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 data governance, getting more value out of enterprise data, and high-speed backup of the data warehouse.

Q&A WITH THE EXPERTS Our Q&A section provides answers from these same experts to the questions they hear most often, complemented by insight from an independent consultant.

FEATURES Volume 22 features two articles from TDWI's director of research and services, Wayne Eckerson. In "Is Your BI Environment Protected?" he asks an important question: what is the biggest threat to your BI environment? In "The Real Value of BI Search," he explains why current trends in BI search may be symptoms of a larger problem.

TDWI RESEARCH *What Works* includes excerpts from TDWI's recent best practices reports: *Deploying Dashboards and Scorecards* by Wayne Eckerson and *Master Data Management: Consensus-Driven Data Definitions for Cross-Application Consistency*, TDWI's latest report from Philip Russom.

BEST PRACTICES AWARDS This section features the winners from TDWI's annual contest, which recognizes organizations for developing and implementing world-class business intelligence and data warehousing solutions.

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

November 2006

tdwi
THE DATA WAREHOUSING INSTITUTE

1201 Monster Road SW, Ste. 250
Renton, WA 98057
425.277.9126 Fax: 425.687.2842
info@tdwi.org, www.tdwi.org

Editorial Director	Denelle Hanlon dhanlon@tdwi.org
Production Editor	Jennifer Agee jagee@tdwi.org
Graphic Designer	Bill Grimmer bgrimmer@tdwi.org
V.P., Bus. Tech. Group & New Bus. Dev.	Ellen Hobbs ehobbs@tdwi.org
Director of Research & Services	Wayne Eckerson weckerson@tdwi.org
Director of Marketing	Michelle Johnson mjohnson@tdwi.org
Production Manager	Marie McFarland mmcFarland@tdwi.org
Art Director	Deirdre Hoffman dhoffman@tdwi.org

1105 MEDIA

President & CEO	Neal Vitale nvitale@1105media.com
CFO	Richard Vitale rvitale@1105media.com
Executive VP	Michael J. Valenti mvalenti@1105media.com
Director of IT	Erik Lindgren elindgren@1105media.com
Director of Circulation & Data Services	Abraham Langer alanger@1105media.com
Director of Web Operations	Marlin Mowatt mmowatt@1105media.com
Director of Print Production	Mary Ann Paniccia mpaniccia@1105media.com
Controller	Janice Ryan jryan@1105media.com
Director of Finance	Paul Weinberger pweinberger@1105media.com
Chairman of the Board	Jeffrey S. Klein jklein@1105media.com

Sponsorship Opportunities: Denelle Hanlon, 425.277.9130, dhanlon@tdwi.org.

Reprints: For editorial and advertising reprints of 100 copies or more, and digital (Web-based) reprints, contact PARS International at 1105reprints@parsintl.com or 212.221.9595, or visit www.magreprints.com/quickquote.asp

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; mail@worldata.com; www.worldata.com

Corporate Headquarters: 9121 Oakdale Ave. Suite 101, Chatsworth, CA 91311; www.1105media.com.

© 2006 by 1105 Media, Inc. All rights reserved. Reproductions in whole or part prohibited except by written permission. Mail requests to "Permissions Editor," c/o *What Works: Best Practices in Business Intelligence and Data Warehousing*, 1201 Monster Road SW, Ste. 250, Renton, WA 98057. The information in this magazine has not undergone any formal testing by 1105 Media, Inc., and is distributed without any warranty expressed or implied. Implementation or use of any information contained herein is the reader's sole responsibility. While the information has been reviewed for accuracy, there is no guarantee that the same or similar results may be achieved in all environments. Technical inaccuracies may result from printing errors, new developments in the industry, and/or changes or enhancements to either hardware or software components. Printed in the USA.

TDWI is a trademark of 1105 Media, Inc. Other product and company names mentioned herein may be trademarks and/or registered trademarks of their respective companies.

TDWI is a division of 1105 Media, Inc., based in Chatsworth, California.

What Works

BEST PRACTICES IN BUSINESS INTELLIGENCE
AND DATA WAREHOUSING

 **CLICK HERE** To request free white papers

CASE STUDIES

Enhancing the Customer Experience and Improving Retention Using Powerful Data Warehouse Appliances	5
Cancer Care Ontario: Creating One-Stop Access to Cancer Services Information	6
RLP Technologies Fuels Better Data Quality through Real-Time Services	7
IXIS Accelerates P&L Analysis	8
Making the Grade with EIM	9
Process-centric BI Improves Collaboration and Delivers Impressive Productivity Benefits	10
Good Technology Boosts Transaction Processing Application by 93 Percent	11
Serving the Public More Efficiently	16
Mentor Graphics Replaces 1,200 Spreadsheets with a Single Web-Based Solution	17
BP Exploration and Production Migrates to a New Business Intelligence Environment	18
Alaska Air Group Soars in Meeting the Needs of More than 17 Million Customers Annually	19
Nationwide Insurance Transforms the Business with Innovative and Award-Winning Finance Data Management Initiative	20

LESSONS FROM THE EXPERTS

The Five-Step Approach to More Valuable Enterprise Data	26
High-Speed Backup and Recovery of Large-Scale Data Warehouses	27
Who, What, When, and WHERE	28
Process-centric BI: Improving Business Processes with Comprehensive SOA	29
Key Requirements for Enterprise Business Intelligence	30
There's More to Business Intelligence than Just Software	31
Company Directory Information and Customer Records: The Schibsted Case	32
How to Launch a Data Governance Program: Practical Guidelines for Technology Implementers	34
It's Not About the Report—It's About What's In It!	35
Intelligent Data Warehousing with ETPL	36
Solving Old Problems with EIM	42
As Business Intelligence Becomes a Strategic Imperative, Is Your Infrastructure Ready to Handle It?	43
High-Performance Technologies Add Up to Successful Business Intelligence	44
Enterprise Information Integration: A Technology for Providing Integrated Views	46
Appliances—Data Mart or Enterprise Data Warehouse?	47

RESOURCES FOR BI & DW PROFESSIONALS

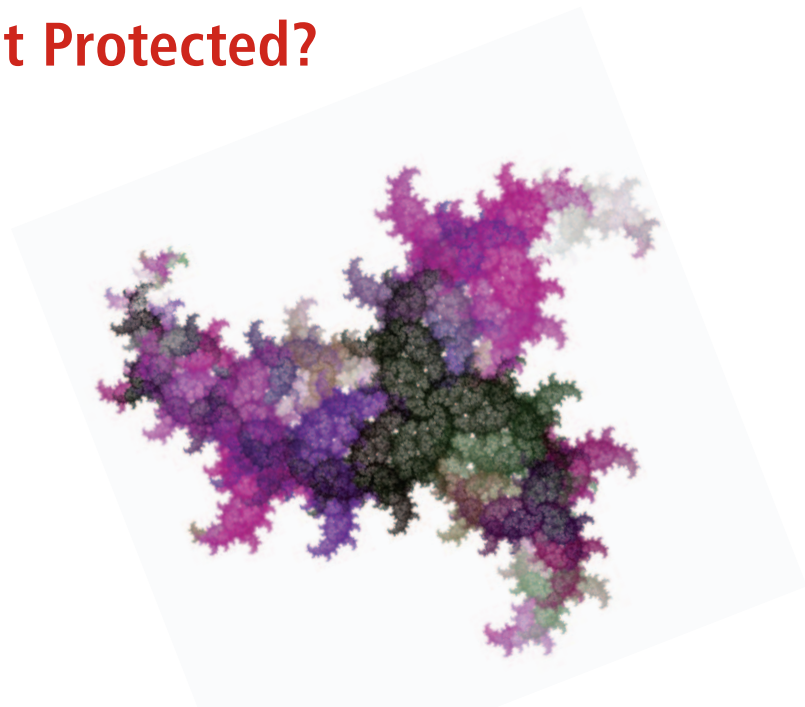
Is Your BI Environment Protected?	2
Q&A with the Experts	12, 37
Deploying Dashboards and Scorecards (Report Excerpt)	22
The Real Value of BI Search	40
Master Data Management: Consensus-Driven Data Definitions for Cross-Application Consistency (Report Excerpt)	48
Winners: Best Practices Awards 2006	52
Directory of Solution Providers	57
About TDWI	62
TDWI Partner Members	64

CASE STUDIES BY INDUSTRY

EDUCATION	
Université du Québec à Montréal	9
ELECTRONICS HARDWARE/SOFTWARE	
Mentor Graphics	17
UGS PLM Solutions	10
FINANCIAL SERVICES/INSURANCE	
IXIS CIB	8
Nationwide Insurance	20
GOVERNMENT	
City of Troy	16
HEALTHCARE	
Cancer Care Ontario	6
INFORMATION SERVICES	
RLP Technologies	7
MANUFACTURING	
BP Exploration & Production	18
RETAIL	
Ahold	5
TELECOMMUNICATIONS	
Good Technology	11
TRAVEL/TRANSPORTATION	
Alaska Air Group	19

Is Your BI Environment Protected?

By Wayne W. Eckerson,
Director of Research and Services, TDWI



Disaster on the Loose

Given the sorry state of our environment and geopolitical relations, I am forced to ask a simple question: When was the last time you reviewed and tested the disaster recovery plan for your data warehouse? Do you even *have* a disaster recovery plan for your data warehouse?

Global warming, by raising ocean temperatures, is spawning more intense hurricanes that pack the power of several nuclear bombs, as witnessed by the devastation wrought by Hurricane Katrina. Global warming also increases the volatility of weather, contributing to more tornadoes and violent thunderstorms that can damage infrastructure and knock out power. And rising temperatures are creating longer and more intense heat waves that strain power supplies, cause rolling blackouts and outages, and spark virtually uncontrollable forest fires that threaten residential and commercial property.

Moreover, scientists say it's only a matter of time before a pandemic breaks loose with the potential to kill millions of people and force the rest of us to work in isolation or quarantine. And the rapidly expanding industrial economies of China and India make the supply and price of oil vulnerable to persistent geopolitical

tensions in the Middle East, unstable governments in Africa, and strained relations elsewhere. A dramatic rise in the price of oil caused by a geopolitical meltdown in any oil-producing nation would quickly send our economy into a tailspin and potentially cause energy slowdowns or blackouts.

Had enough? While natural and geopolitical disasters are on the rise, they aren't the biggest threats to your BI environment. According to research by *Information Age*, a leading UK magazine for executives, most IT executives believe the greatest threats to the continuity of their IT operations are internal system failure (65 percent) and viruses (45 percent). Natural disasters registered at 32 percent and power and communications outages at 33 percent.

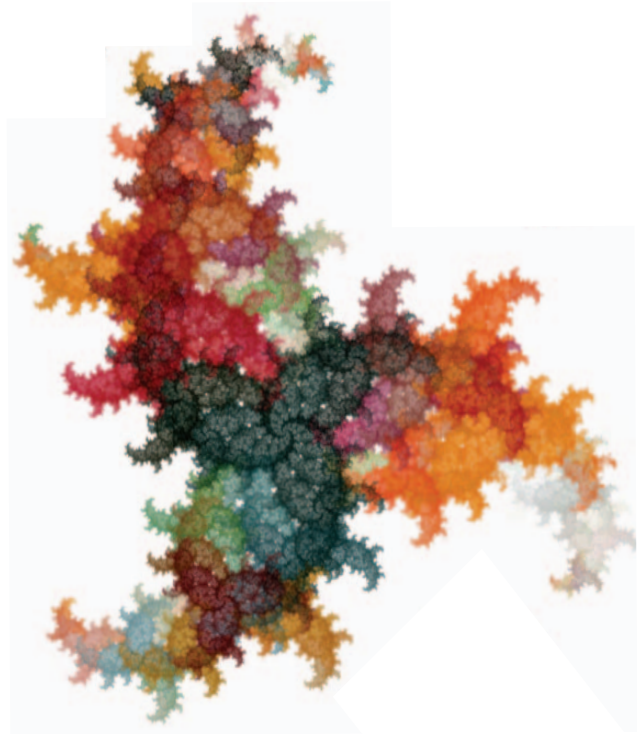
Role Reversal

Ten years ago, there was little need to create a disaster recovery plan for data warehouses and the reports and applications they support. At that time, the vast majority of data warehouses were loaded in batch on a monthly basis from a half-dozen or so source systems. Most loads were fairly small, and even the biggest data warehouses were less than a couple

of hundred gigabytes in size. Not surprisingly, most data warehousing teams didn't have a disaster recovery plan, let alone a backup strategy. The common sentiment back then was that if the data warehouse crashed, you could simply refresh the data warehouse in its entirety from source systems once everything came back online.

Today, most data warehouses have become mission-critical systems. Many now capture and update transactions on a real-time basis and support dozens of run-the-business applications. Business users have become so dependent on data warehousing information to make daily business decisions that they practically crucify data warehousing managers if the system goes offline even for a few hours. Moreover, as a decision-making engine, a data warehouse is critical to helping organizations respond in an optimal fashion when disaster strikes. Data warehouse reports can help executives figure out how to prioritize activities, allocate resources, and reassign staff to deal with emergencies. A decade ago, the data warehouse may have been the last system restored after a disaster; today, it should be the first system to come online in an emergency.

A decade ago, the data warehouse may have been the last system restored after a disaster; today, it should be the first system to come online in an emergency.



How Protected Are You?

Research shows that a majority of organizations are confident in the resiliency of their IT systems. Most have disaster recovery plans that safeguard the business from short- and long-term disruptions. Maybe the disaster recovery plan even includes the data warehouse, the servers it runs on, and the reports and applications it supports. Since many data warehouses now run within corporate data centers governed by corporate IT policies that include business continuity and disaster recovery planning, it is a good bet your organization has insured its data warehousing assets to some degree.

Unfortunately, most disaster recovery plans don't go far enough to protect an organization against costly disruptions. Disaster recovery planning is insurance, and most companies insure only what they can afford—not what they need.

Has your organization prioritized the business processes and applications that are critical to its operations? If the data warehouse is a top priority, what about the ETL engines that populate the data warehouse and the BI servers that generate and distribute critical reports? A chain is only as strong as its weakest link, and a data warehouse is a complex environment

comprising multiple systems, applications, and interdependencies with internal and external systems. The data warehousing environment can't be fully restored until every one of its components is brought back online.

When was the last time you really tested the disaster recovery plan for your data warehouse? If you practiced recovering from a database failure, you only completed part of the test. You need to restore clients, servers, networks, storage, applications, and databases to fully simulate a recovery situation. And if you conducted your tests a year ago, there's a good chance that your plans are out of date. Since a data warehouse is an adaptable system, it is constantly changing to answer new questions that business people ask. So, the queries, reports, metadata, ETL workflows, aggregates, and so on have probably changed since your previous test. Moreover, the questions that business people ask during an emergency may be very different from what they normally ask.

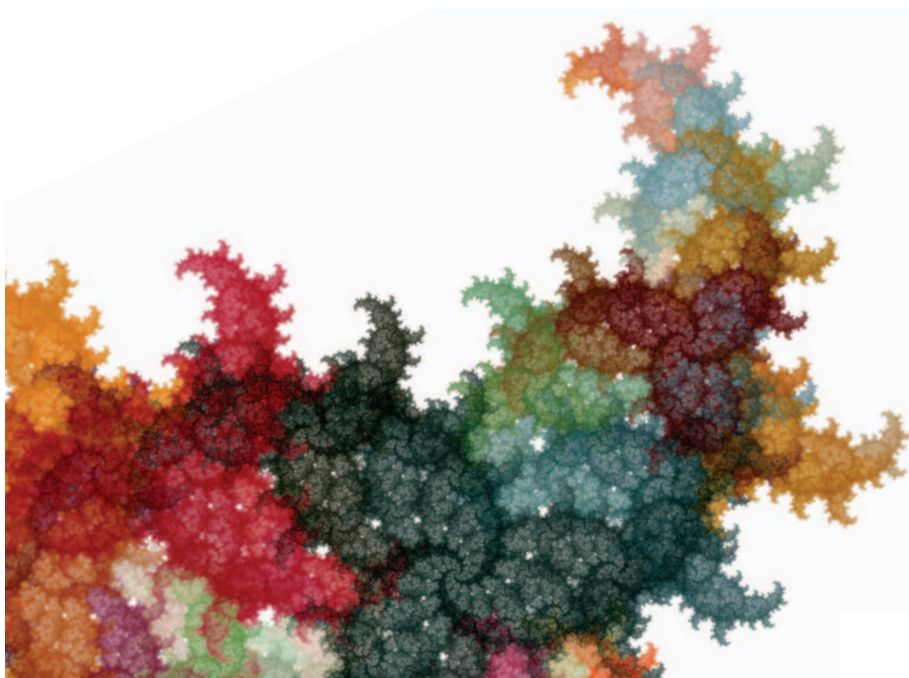
The key to resiliency is not just flexible, redundant systems; it's people. During a disaster, there is a lot of chaos and confusion. Many key personnel may be absent or unable to work or access systems. Thus,

you need redundancy not only in your systems, but also in your staff assignments. Your team members should all be schooled in what to do in a variety of emergency situations—and be ready to play multiple roles as needed.

Be Prepared!

Disaster recovery puts a premium on good-quality, up-to-date, end-to-end metadata, something that few organizations have successfully implemented. Metadata is critical for performing impact assessments—when something in a source system changes, you need to know how it will impact every other component in the system, down to metrics within end user reports. In an emergency situation, data warehousing teams can be seriously hamstrung in their ability to meet recovery time objectives (i.e. time to recover business functions), critical data points (the point in time from which data must be recovered), and recovery point objectives (time to recover data) without access to a dynamic, comprehensive metadata management system.

Of course, the data is the heart and soul of a data warehousing environment, and organizations must devise a good strategy for safeguarding data against



power failures, network outages, floods, storms, or other disasters. Most organizations perform backups to low-cost tape, which is shipped and stored offsite. While it takes a long time to recover a data warehouse from tape, most of this data is historical and doesn't have high value during an emergency. To protect more recent information, organizations should replicate or snapshot data as it moves through the ETL process and store it on disk in a disaster recovery system, which archives or deletes the data after an appropriate period of time, usually a few days or weeks.

Most data warehousing teams understand the need to manage the lifecycle of data warehousing information. Unfortunately, these teams often don't anticipate disaster striking twice. Ideally, the online backup system should be maintained offsite so a data center problem doesn't disrupt both the primary and backup systems. (This obviously is more costly and requires high-speed network connections.) They also don't have a backup for the backup if the offsite system goes down. Most also don't envision a disaster lasting more than a few days. Given that many businesses are *still* not fully functional in the wake of Hurricane Katrina,

we need to extend the duration for how long we expect disasters to last. Finally, many offsite backup systems don't protect companies from viruses, which propagate internally. An offsite system should have an internal gate that delays real-time propagation by several hours to safeguard against software attacks.

Conclusion

It's not fun being the voice of gloom and doom, and no one wants to spend money to avert something that may never happen. But it seems to me that we are witnessing an inflection point in the number of crises, disasters, and geopolitical tensions caused by environmental degradation and political polarization. Of course I may just be hypersensitive to world and global events (or just tuning in for the first time in my life). Nonetheless, there is nothing like word of a possible impending crisis to impel us to dust off our disaster recovery plans. It's better than waiting for a real-life disaster to test the effectiveness of our plans. ■

Wayne W. Eckerson is the director of research and services for TDWI. He has 17 years of industry experience and has covered data warehousing and business intelligence since 1995. Eckerson is the author of many in-depth reports, a columnist for several business and technology magazines, and a noted speaker and consultant. His book, Performance Dashboards: Measuring, Monitoring, and Managing Your Business, was published by John Wiley & Sons in 2005. He can be reached at weckerson@tdwi.org.

Enhancing the Customer Experience and Improving Retention Using Powerful Data Warehouse Appliances

Commentary by Brent Buelo, Manager of Data Warehouse Services, Ahold

Customer Profile

Ahold operates 800 supermarkets in the U.S. under brands such as Stop & Shop, Giant Food, and Tops Markets, and is a division of Netherlands-based Royal Ahold, the world's third-largest food retailer. Ahold performs advanced analytics for crucial insights into shoppers' buying patterns, and views its customer data warehouse (CDW) as central to the way it manages product availability and ensures customer satisfaction.

Business Challenge

Ahold's legacy data warehouse system could not keep pace with growing data volumes and query complexity. Essential reports often took days to execute, making drill-down analysis impractical, and many ad hoc requests were impossible to fulfill. With data growing exponentially, Ahold needed a system able to query terabytes of sales and operational data, and deliver fast, comprehensive results.

Netezza Architecture Breaks Down Previous Barriers

Ahold chose the Netezza Performance Server data warehouse appliance because of its ability to analyze millions of customer purchases faster and with greater flexibility than any other option in the market. The NPS system integrates database, server, and storage into one system designed specifically for high-volume, complex analytics at a low total cost. The Netezza architecture also eliminates the need for tuning, indexing, data partitioning, or aggregation required by other systems.

"We chose the NPS data warehouse appliance because it delivers the increase in performance we wanted without reengineering our entire CDW," explains Al Clevenger, director of data management and data warehouse services, Ahold. "Now we can achieve results in minutes instead of hours, eliminating stale data

and enabling us to be even more flexible and targeted with our decisions. The system was also two-thirds the cost of our previous customer data warehouse, including ongoing support and maintenance."

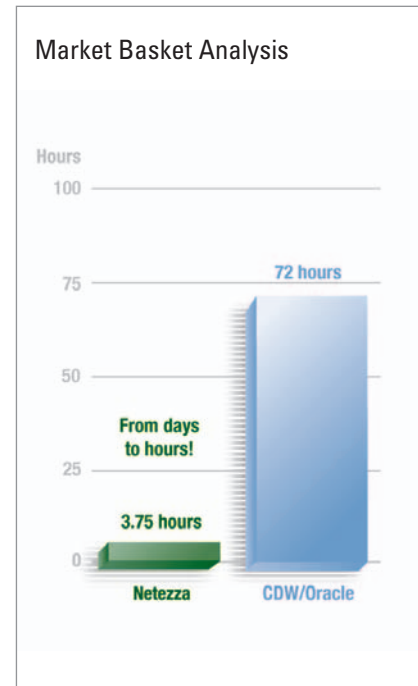
Operational Simplicity, Low TCO

Ahold was able to take advantage of Netezza performance without reengineering its CDW—usually a costly, time-consuming project when migrating to new data warehouse systems. The simplicity of the Netezza appliance made it easy to deploy, integrating seamlessly with existing business intelligence applications and ETL infrastructure. The low purchase price is complemented by minimal ongoing maintenance, enabling Ahold to reduce DBA commitment from more than two full-time staff members to less than half of one person's time.

Results: Digging Deeper, Working Faster, Understanding More

By boosting performance by orders of magnitude, the Netezza architecture allows Ahold's merchandising teams to perform much faster, more sophisticated analyses of customer buying patterns. For example, one complex query that previously took 72 hours was completed in four hours using the Netezza appliance. Ahold is now running 133 percent as many analyst reports by nearly 20 percent more users, and overall productivity has increased so far by 25 percent.

The system's speed and simplicity also allows Ahold to go much deeper into its data with more iterative analyses, and to understand customers in a much more comprehensive way. Users are no longer constrained by system limitations, and are realizing what can be gained from their data when results return in seconds or minutes rather than hours, days, or sometimes never. As a result, Ahold is



able to use in-depth analysis of purchasing trends to continually enhance the customer experience, with optimized store layouts, closely targeted sales promotions, and streamlined inventory management.

With Netezza, Ahold Now Understands its Customers in a Much More Comprehensive Way

By replacing its traditional data warehouse with the Netezza Performance Server data warehouse appliance, Ahold has a data warehouse system with the processing power and flexibility to fully leverage the detailed data it collects about its customers. With query times reduced dramatically, Ahold is able to continually innovate and fine-tune its analyses by asking complex questions that were previously out of reach. The result: better-informed decisions and a leadership position in the food retail industry—with increased sales, lower costs, and greater customer retention. ■

Cancer Care Ontario: Creating One-Stop Access to Cancer Services Information

Commentary by Richard Duck, Portfolio Manager, Data Warehousing, Cancer Care Ontario

Cancer Care Ontario is the Ontario government's principal adviser on cancer care in the province. Cancer Care Ontario's (CCO) primary mission is to provide the Ministry of Health in Ontario with funding recommendations for the delivery of cancer services to the 14 different health regions in Ontario. To facilitate this, CCO collects data from a variety of institutions and hospitals. The data collected ranges from cancer diagnosis, pathology data, and surgical data, to hospital in- and outpatient data, as well as a variety of other program-specific information. This information enables CCO to help guide the cancer system in the province, as key health informatics resources plan and manage for the burden of cancer today and in the future.

An Accessible, Easy-to-Use, Web-Based Portal

In September 2005, Cancer Care Ontario launched its first implementation of the data warehouse and a new, Web-based, analytic tool called iPort. To date, iPort has been rolled out to over 250 cancer planners in Ontario and several other healthcare providers, public health units, administrative functional groups, and the ministry itself. iPort enables its users to view prepackaged cancer reports easily and without any detailed analytics in a section of the tool called "quick facts." Quick facts enable high-level users to log into iPort and get an easy-to-follow overview of the cancer system by subject area, such as cancer surveillance or cancer activity. In each of these areas, quick facts tell users the story of what is happening in the province, and how they can use these reports to gain a better understanding of the critical information provided. In addition, iPort provides more frequent users with the capacity to build their own "prompted" reports in a section of the tool called "detailed analysis." This section

enables users to select the attributes, metrics, and filters they would like to see in their reports and build them to their own specifications. Both of these sections have proved highly useful for CCO users as they work to plan and manage the cancer system.

iPort and the underlying data warehouse have been built using several technologies, including MicroStrategy 8 (recently upgraded to 8.02). The capabilities of MicroStrategy and the creation of iPort have helped to change the external perception of CCO from a slow, process-heavy, silo-driven agency to a highly effective information management organization that empowers users to get the data they need quickly and easily.

Creating and implementing iPort has facilitated robust reporting and analysis capabilities via an accessible, easy-to-use, Web-based portal. The solution provides increased efficiency, improved access to information, and increased reporting and analysis on key clinical, research, financial, planning, and measurement questions. It has also improved productivity and reduced labor costs.

iPort's Relevance to Other Healthcare Organizations

Not only has the introduction of iPort and the addition of the data warehouse had a positive impact on the way that CCO gathers, stores, manages, and disseminates information, but in addition, several other healthcare organizations are now approaching CCO for assistance with their data warehouse road maps and other similar tools. To date, iPort has been introduced and demonstrated at more than 25 internal and external executive- and ministry-hosted events, as well as data warehouse industry events.

In 2006, CCO will start using the data warehouse and iPort to create the first fully integrated cancer continuum



reporting tool, by patient, in Canada. This function will support a highly advanced level of reporting and will enable cancer planners and researchers to track patients from population-based risk factors and screening, through diagnosis and treatment, to outcome.

iPort Stands Out and Delivers a One-Stop User Experience

CCO information consumers now have "one-stop shopping" for cancer services information. iPort was designed by users, for users, and delivers an appropriately organized and integrated information architecture targeted at specific users and their needs. iPort presents consistent answers to important business questions, removing subjective interpretations and providing an authoritative source on provincial cancer information. In doing so, iPort has garnered strong executive sponsorship, business sponsorship, and cancer planner/manager support; strong buy-in from the user community; high user adoption; and increased efficiency in cancer planning results.

In short, the data warehouse and iPort, powered by MicroStrategy, enables CCO to deliver information products in the most effective manner, with unlimited possibilities for improvement and growth, and establishes a strong component in CCO's continuing effort to reduce the burden of cancer on Ontarians. ■

RLP Technologies Fuels Better Data Quality through Real-Time Services

Commentary by Daniel Teachey, Director of Corporate Communications, DataFlux Corporation

The Business

RLP Technologies is a subsidiary of R.L. Polk & Co., a leading provider of automotive information and marketing solutions. RLP Technologies specializes in building industry-leading data solutions that serve as the foundation for focused, in-depth research and analysis across multiple industries—enabling these tools and applications to “mine” intelligence from the data.

The Challenge

R.L. Polk has set the standard for automotive vehicle and consumer data for many years, continually improving its data management methods along the way. But Polk needed to move beyond incremental improvements and develop a new, innovative approach that would greatly enhance Polk’s existing core foundational data warehouse.

As a result, RLP Technologies was asked to initiate a reengineering effort to implement breakthrough technologies that would collect, standardize, and enhance data from disparate sources, and compile them into a “single source of the truth” for distribution to analytical and operational applications. The project led to the development of the Enterprise Information Factory (EIF) as a comprehensive, integrated application for enterprise information management. As a key part of the reengineering initiative, RLP Technologies examined several data quality vendors to validate the interoperability of the solution within their target architecture—and assess the performance of these products when managing data with even the most complex business requirements.

The DataFlux Solution

RLP Technologies selected DataFlux dfPower Studio to analyze, improve, and control data quality within the Enterprise Information Factory that was developed

RLP Technologies was asked to ... collect, standardize, and enhance data from disparate sources, and compile them into a “single source of the truth.”

to address R.L. Polk’s data requirements. At the outset of the project, data profiling provided an in-depth data discovery phase to catch data quality issues as soon as possible in the data’s lifecycle.

The DataFlux tools used to perform name-and-address standardization included a business rules builder and the Quality Knowledge Base (QKB) from DataFlux’s dfPower Studio. These elements gave Polk the flexibility to create custom rules that would easily integrate with all future releases of the DataFlux software.

The DataFlux Integration Server allowed RLP Technologies to expose business rules for data quality via a service-oriented architecture (SOA), creating a real-time processing architecture to ensure that data entering RLP Technologies’ solution met Polk corporate standards for data integrity. Once completed, the project encapsulated decades of unique and complex business rules that parse and standardize name and address data from vehicle title and vehicle registration records.

“DataFlux’s powerful data quality engine provided us with all the feedback necessary for our Enterprise Information Factory to evaluate rules, measure thresholds, and adjust for violations,” said Joe Lafeir, vice president of product development for RLP Technologies. “DataFlux’s architecture and flexibility provided a means to tailor the name-and-

address standardization solution with the complex set of business rules that we needed to move from the legacy mainframe environment.”

The Results

Thanks to DataFlux’s data quality integration technology, RLP Technologies has a more consistent, accurate, and reliable foundation of data. The reengineering project served as an incubator for RLP Technologies to offer Polk’s innovative software solution commercially, allowing others to benefit from higher-quality data.

With the help of DataFlux, RLP Technologies created an engine that enhances the timeliness, accuracy, and quality of foundational data for Polk’s analytical and operational applications. Polk’s applications enable the automotive industry to analyze the market to make critical decisions about their businesses, communicate more effectively with prospects and customers, and evaluate the effectiveness of their programs.

“DataFlux provided us with excellent support throughout the entire process,” said Lafeir. “The technical support and the online customer portal kept us updated about technology advances that we could embrace. We engaged DataFlux’s professional services early in the project, and they provided deep content knowledge to help get the project off the ground.” ■

IXIS Accelerates P&L Analysis

Commentary by Damien Le Marchand, Project Manager, IXIS CIB

For the investment bank IXIS CIB, a subsidiary of the third largest banking group in France, the process of calculating its profit and loss (P&L) on a daily basis involves analyzing 300 million transaction lines.

Until recently, IXIS CIB used a reporting system designed solely for trading floor operators. This system relied on both a production database and an information database. Both of these were transactional management databases.

“We carried out batch loading from the production database to the information database. Requests from traders were made alongside this transaction processing, which could cause performance problems. With the growing complexity of analysis required by traders, loading times increased to four hours. Due to these constraints, background history was gradually reduced to a period of no more than six months in the production database and 15 consecutive days in the information database,” explains Damien Le Marchand, project manager at IXIS CIB. The huge volume of processing together with peak loadings caused bottlenecks and affected response times. Complex requests could take up to three hours.

To better support the activities of its risk managers, the bank realized it needed a more efficient and scalable system—a system that would enable it to reconcile P&L and accounts. The new system would have to combine P&L lines, the dynamic risk parameters in each portfolio, and cash flow, which meant the target system would need the capacity to handle much larger data volumes and a constantly increasing number of requests.

Evaluating Alternatives

IXIS CIB identified two ways to address this challenge. One involved reviewing working patterns and changing operational procedures. The second involved

consolidating all transaction data into a central information database, isolating reporting and analysis from transaction processing. The key to making this second method work was the ability to deploy a robust, high-performance data management system with the capacity to scale as data storage requirements increased, and to expedite loading and response times regardless of the number of users and the complexity of requests.

IXIS CIB chose the second way and found its data management solution in the Sybase IQ analytics server. After achieving positive results in testing, the bank began to integrate P&L lines and the dynamic performance indicators coming from front-office trading applications, namely Summit for fixed-income and foreign-exchange business, and Calypso for credit derivatives and complex rates, in total representing 300 million lines.

The project went live within a month thanks to support from Sybase Professional Services, which facilitated the porting of the previous central information database to Sybase IQ and the transfer of skills to the P&L IT team.

A future phase will integrate two other front-office applications: Sophis for shares and ACBS for financial accounting and cash flow lines. All data will be available through Business Objects reporting tools via an intranet.

Sybase IQ Delivers in Real Time

Thanks to Sybase IQ, transactions from the previous day are available to trading operators and risk managers before the markets open, with load times reduced to 30 minutes. Sybase IQ provides risk managers with both the speed and the detail to allow them to complete their tasks successfully by zooming right in on the minute details of each transaction. This has all been achieved without impacting database availability or performance.

“By separating out the transaction processing database, we’ve achieved response times that are between 10 and 20 times faster than our earlier system.”

—Damien Le Marchand, IXIS CIB

“By separating out the transaction processing database, we’ve achieved response times that are between 10 and 20 times faster than our earlier system. Responses do not exceed four minutes, even for the most complex cases,” says Le Marchand. The speed of processing calculations that are linked with the central information base, which consolidates all transactions and cash flow lines, now provides full visibility of trading results the following day. Moreover, daily consolidation gives the accounts department access to data following the close of each day’s trading, allowing it to meet the reduced 15-day reporting cycle required to comply with the IFRS standard.

The bank has also realized IT cost reductions, which by themselves constitute an impressive ROI. ■

Making the Grade with EIM

Commentary by Lise Carriere, Director, Institutional Research,
Université du Québec à Montréal

Consolidating Student Lifecycle Data From Disparate Sources

Université du Québec à Montréal (UQAM) offers a wide range of programs to approximately 40,000 undergraduate, graduate, and research students. It is a public, French-language urban university, located in Montreal, Canada.

UQAM closely monitors the lifecycle of each of its enrolled students from initial application to graduation. It's critical that the university understand trends in its student population—such as demographics, attrition rates, and program-specific information—in order to continue attracting first-rate students and professors.

With tens of thousands of current and potential students each year, maintaining data in student lifecycle files is no small task. Historically, this data was stored in a number of independent systems—meaning that UQAM could not provide a single, coherent view of its data. Users received different results depending on which system they used. “It was quite a mess,” says Lise Carriere, director of institutional research at UQAM. “There was a huge need to make data accurate and accessible to users and administrators so they could make better decisions.” In addition, because the data was stored in disparate systems, UQAM had a difficult time tracing database errors.

Seamless Data Integration For Quick Access

UQAM needed to consolidate all of this data—contained in disparate systems—without complex coding and script writing. After some investigation, the university selected and deployed BusinessObjects Data Integrator, a Business Objects enterprise information management (EIM) tool. Data Integrator allows the university to bring together multiple sources of data and gain access to accurate, timely, and trustworthy information. It also facilitates

greater system analysis because it provides documentation of the data integration processes. “Documentation is always the last task to be completed, if ever,” notes Carriere. “Having this content readily available is especially reassuring to the university’s small team, which has to deliver results in a short time period.”

By implementing Data Integrator, end users across the university now have a single, integrated view of their institution’s business. Data contained in UQAM’s operational systems is loaded seamlessly into the data warehouse without complex coding.

Reduced Time and Cost, Boosted Data Quality

Thanks in part to the relationship built with Business Objects and the firm’s EIM offerings, UQAM has realized some important benefits across the entire university. The time saved in development, especially from not having to write custom scripts, was key, according to Carriere. “The project as a whole would not have gone forward, given our small team,” she says. “Data Integrator made the project feasible in a reasonable amount of time.”

The university was also able to more quickly integrate additional sources into its data warehouse. This, in turn, allowed more time to implement its BI programs. “We’ve been able to minimize IT infrastructure costs since the deployment of the project,” says Carriere. “We’ve greatly benefited from the reusable development efforts and from the greater level of data quality—since the same jobs and scripts are consistently used and easily updated if there are any changes in the database.”

Djaffar Ben Meddour, senior analyst of institutional research at UQAM, notes: “Since we track the progress of our students, data quality is critical to ensuring we’re working with accurate information. We use the whole stack of Business



Objects products to provide users with a single, integrated view of our students and analyze trends in our student population. EIM ensures that we have trustworthy information to drive our BI and performance management programs.”

Carriere is relieved that the deployment has satisfied the overwhelming requests for data insight from university employees. “Finally,” she says, “we can access data that we desperately needed to see. And we can get it faster than ever before.” ■

“Finally, we can access data that we desperately needed to see. And we can get it faster than ever before.”

—Lise Carriere, Université du Québec à Montréal

Process-centric BI Improves Collaboration and Delivers Impressive Productivity Benefits

Commentary by W. Michael White, Director of Financial Planning and Reporting, and Eric Kline, Director of Enterprise Applications Operations, UGS PLM Solutions

The Challenge

UGS PLM Solutions (an EDS company), a rapidly growing product lifecycle management software developer, found itself burdened with an inefficient and complex business reporting system consisting of Microsoft Excel to track business performance; SAP R/3 for financial transactions, sales, distribution, and human resources; and Hyperion Pillar for annual financial planning and forecasting.

As the company grew, so did the complications of maintaining such a system. The unsophisticated system restricted IT to generating only monthly reports, and the finance department was dissatisfied because the reports were limited and not in real time. Moreover, the system lacked proper security integration.

The finance and IT departments jointly outlined requirements for a solution that would improve decision support by enhancing the enterprise management reporting and analysis systems. In addition to reducing monthly report generation processing time, replacing unsupported legacy systems, and improving flexibility, the solution would also need to increase user friendliness and speed of access by offering Web-based data delivery.

The Solution

Ultimately, arcplan emerged as the optimal BI solution for UGS. The ability to access any and all data sources enables arcplan customers to develop interactive BI applications in record time (installation time averages less than three months) and cost effectively by avoiding costly data migrations. W. Michael White, director of financial planning and reporting, commented, "Other applications would have required us to develop an overlay security system, thus adding to the project cost and time. arcplan Enterprise provides boardroom-quality reports. It empowers the end user to slice, dice, drill down,

"Our arcplan deployment dramatically increased business user, analyst, and developer productivity. It demonstrated that we could be flexible even with massive increases in data."

—W. Michael White, UGS PLM Solutions

drill up, and drill through all types of data sources."

The ability to mimic existing systems led to rapid user adoption. "If they can use a browser, they can understand the logic behind arcplan," White said of the business users. "We purchased arcplan in March, integrated it with Hyperion in May, and by early July, we were running the first arcplan apps. Financial 'super users' quickly became fluent in arcplan, and in two weeks were able to replicate the original Excel reporting structure," White recalls. "The beauty of arcplan is that the finance staff can be the ones to develop the appropriate end-user reports without having to go to IT for development resources."

The Benefits

Six months after go-live, system resources faced a major test—a merger. UGS PLM Solutions is the combination, through acquisition, of SDRC and Unigraphics. "Here's where arcplan's flexibility really shone, and it's probably the biggest advantage we've realized, although we couldn't have predicted these circumstances a year earlier," said White. The reporting templates were implemented within two months after the organization and systems were consolidated.

arcplan enables organizations to roll out process-centric BI analytics in less than one-third of the time of other BI deployments and against existing, disparate data sources, improving decision making and reducing ongoing IT costs. "Our arcplan deployment dramatically increased business user, analyst, and developer productivity. It demonstrated that we could be flexible even with massive increases in data. And it clearly succeeded in helping us get information quickly to our executives," said White. "We identified that over a period of five years, the Web deployment would create over \$500,000 in true system cost savings," he continued.

Eric Kline, director of enterprise applications operations, explained IT's perspective. "The combination of arcplan and Essbase allows for a dramatic improvement in the user-friendliness of our reporting environments. It further enhances the access to multiple dimensions of data in various formats from an end user perspective. arcplan's report development flexibility and ease has reduced the dependency of IT personnel from having to write custom SAP performance measurement reports."

Looking ahead, White states, "We haven't even begun to scratch the surface of arcplan's potential." ■

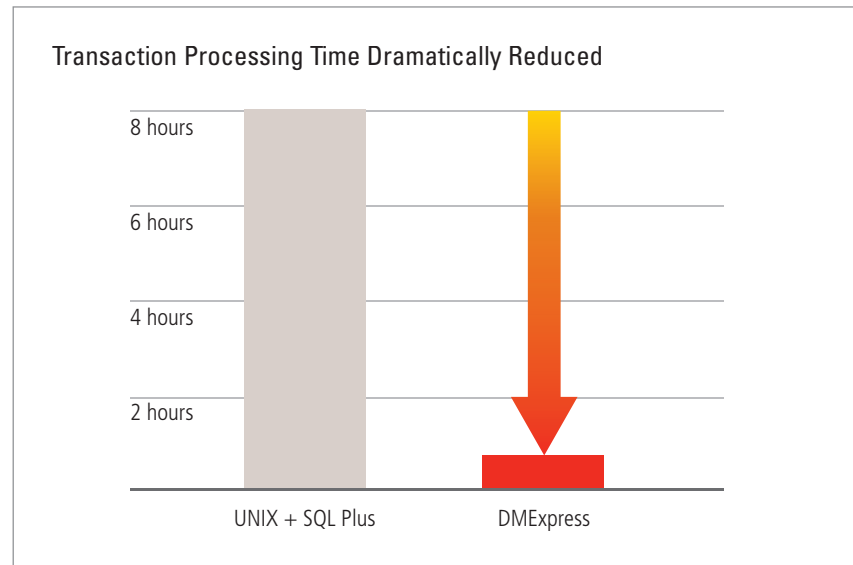
Good Technology Boosts Transaction Processing Application by 93 Percent

Commentary by Simon Ha, Database Administrator, Good Technology

The Challenge

California-based Good Technology is the industry leader in enterprise systems for corporate message and mobile application access. They offer corporate customers the ability to access various mission-critical enterprise applications via PDAs and other handheld devices. Mobile workforce employees can access Microsoft Office, Oracle, salesforce.com, and Siebel through wireless networks.

When funneling transaction usage data from a variety of live collection points, however, Good Technology ran into a processing bottleneck. This transaction data—which is similar to an individual’s cell phone usage—was needed for analysis by Good Technology’s accounting and marketing departments. They needed to



Organizational Profile

Industry leader in enterprise systems for corporate messaging and mobile application access

Business Needs

- Minimize the elapsed time of daily transaction processing application
- Apply various transformations to high volumes of data to allow for simplified analysis

Environment

- SUN 420R with 4 CPUs
- Oracle 9i data warehouse
- UNIX Utilities
- SQL Plus

Benefits

- 93% performance improvement in the processing of data
- Join, append, and lookup operations efficiently transform 35 million rows of transaction data per day

process and analyze the data to extract critical business intelligence. Because the company’s operations were 24/7, raw data was continually streaming in, sometimes at the rate of 35 million rows per day. The problem was that the processing, which was done using SQL Plus and some native UNIX utilities, was taking 16 hours. Good Technology’s staff needed to minimize the elapsed time of daily transaction processing and apply various transformations to the high volumes of data to allow for simplified analysis.

The Solution

Good Technology requested a proof of concept from Syncsort—who in turn committed to a demonstration of a significant elapsed time savings with the implementation of DMExpress on this mission-critical application. Using 18 million rows of actual production data, which would have taken about eight hours’ processing time, the engineers at Good Technology ran a sequence of DMExpress steps and were able to bring the processing time down to just 30

minutes, a reduction of 93 percent. Very pleased with this outcome, Good Technology licensed the Syncsort product.

The Benefits

“Not only were we able to improve the performance of our transaction usage application; we also gained functionality through the use of the DMExpress GUI,” says Simon Ha, database administrator at Good Technology. Join, append, and lookup operations are now easily and regularly applied to up to 10 gigabytes of data per day. Results have been so satisfactory, in fact, that Good Technology is planning to expand the use of DMExpress to other departments within the company. ■

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. Mark Hammond, an independent consultant, provides his analyst viewpoint to each Q&A.

arcplan, Inc.

“Process-centric BI” and “operational BI” are both hot topics these days. Since operations are made up of processes, is there a difference?

Yes, and the difference lies in the level of the activity. With operational BI, vendors are extending BI from the strategic and tactical levels into the operational level, where BI has not been used—until now. Process-centric BI, on the other hand, does not fit neatly within a particular level. Instead, it follows business processes, and the underlying data, through different levels. In this way, process-centric BI can be strategic, tactical, and operational, all in the same application.

Analyst Viewpoint

Another distinction is that operational BI is usually most relevant to hands-on systems such as call centers and business performance management (BPM) or business activity monitoring (BAM), whereas process-centric BI lends itself to intelligently automating such processes as supply chain and pricing based on business rules. Distinctions aside, both operational and process-centric BI share the same objective of reducing data latency from weeks or days to hours or minutes to enable more precise business execution. The service-oriented architecture offers IT organizations a way to non-invasively embed analytics in both operational applications and processes to minimize impact on business-critical systems.

Request more information about arcplan, Inc.

Business Objects

What components make up an EIM technical framework?

Enterprise information management (EIM) is the strategy, practices, and technologies needed to deliver a comprehensive approach to managing data. While many of the components that make up the technical framework of EIM are not new, the category itself and the ability to deliver an integrated approach is new to the market. Data integration (ETL), data quality, metadata management, and data federation are the core components companies implement within an EIM strategy. While organizations may not need to adopt all components or may stagger their implementation, it is essential that the solutions that make up the EIM framework support and integrate easily with each other.

Analyst Viewpoint

If source data is poorly integrated, inconsistent, and unreliable, a glitzy BI front end can amount to little more than whipped cream on a pile of garbage. Most organizations have experience with the technical components of an EIM framework (data integration, federation, quality, and metadata management), and have ad hoc solutions across their information infrastructure. The EIM challenge is to avoid the pile-of-garbage BI data syndrome by integrating these technologies into an EIM framework (driven by strategic objectives and defined practices) to provide a sound information infrastructure upon which BI can thrive. For some EIM adopters, a key choice will be whether to deploy best-of-breed technologies or an integrated solution from a single vendor.

Request more information about Business Objects

DataFlux Corporation

What are the advantages of implementing service-oriented architecture (SOA) in a data quality program?

Many businesses acknowledge the importance of having a unified view of company data across multiple enterprise applications. Data quality assures that data is consistent, accurate, and reliable, and service-oriented architecture (SOA) guarantees the data's integrity as it arrives from multiple sources. SOA provides a rapid time-to-value, allowing you to create a business rule for data quality once—and replicate it across any application. SOA allows business users to more effectively enforce data standards, creating an enterprisewide standard for data quality. Data quality management via Web services in an SOA approach is the best method to ensure overall data integrity.

Analyst Viewpoint

Despite widespread recognition of issues with poor data quality, “dirty data” remains one of the most pernicious problems in enterprise data management. Attempts to improve data quality with isolated or departmental solutions have in many cases resulted in additional siloed solutions that fail to address the issue on a cross-enterprise scale. SOA offers organizations an opportunity to transparently integrate data quality services to cleanse, update, and propagate accurate data across loosely coupled heterogeneous systems for enhanced business value. Because SOA typically involves architectural changes at a foundational level, organizations embarking on an SOA-based data quality initiative will often be best served by an incremental approach.

Request more information about DataFlux Corporation

DATALlegro, Inc.

How do you handle a mixed workload of short- and long-running queries, all competing for system resources?

Current RDBMS platforms allow multiple concurrent connections and queries to be submitted by independent users. The result is a mixture of short-running queries and long-running, process-intensive queries. Without a priority management system, as the number of queries increases, the performance for each individual query degrades almost linearly.

Standard RDBMS vendors address this issue by setting higher priority to specific user accounts or to certain connections, resulting in administrative overhead and dissatisfied end users. The most advanced DW appliances provide automatic workload management without a complicated configuration. Resource priority is automatically and continuously reassigned based on real-time system throughput. The resulting system is one that automatically matches performance to user expectations, even under high load conditions.

Analyst Viewpoint

Many factors must be taken into account when engineering query prioritization and optimization mechanisms, including I/O, CPU speed, memory, and bandwidth. Ultimately, the most important factor is the business priority associated with each query. Today's configuration tools help simplify management of complex query environments with finely grained controls that may be tied to user role, time of day, and other criteria. Slick graphical displays and outlier alerts help administrators monitor the systems. Organizations will benefit from both automated workload management and periodic reviews of query prioritization and performance. In addition, diligent, trial-and-error application of performance tuning techniques with built-in or third-party tools or data accelerators can pay significant dividends.

Request more information about DATALlegro, Inc.

ESRI

We've seen compelling demonstrations of geographic information systems (GIS), but aren't they difficult to learn and implement?

Today's server-based geographic information systems can be used through the interface provided in your business intelligence applications. All of the leading BI vendors, or their partners, have built in the ability to link to GIS analysis and mapping capabilities. In many cases, your BI application will guide you through the process of linking your data to commercially available GIS feature data (maps). The most commonly used GIS feature data is included with the GIS application, or it can be accessed as a Web service.

Analyst Viewpoint

We've seen compelling demonstrations of geographic information systems (GIS), but that, as they say, was then. This is now. Over the past several years, GIS and business intelligence vendors have made significant strides in putting GIS client and Web-based applications on the BI map with ease-of-use enhancements aimed at business users, workflow automation, and standard protocols that enable integration with RDBMS and business applications. Organizations that wish to exploit the power of GIS in such areas as marketing, supply chain, or finance can test the waters with a low-end GIS product (such as Microsoft MapPoint) before moving up to the greater breadth and functionality of GIS specialists such as ESRI.

Request more information about ESRI

FAST

How can you ensure that the data used to make business decisions is reliable?

In order to establish a single version of the truth on which to base corporate decisions, you must prevent errors introduced in transactional systems from propagating to the aggregated information residing in your data warehouse.

This requires operational data cleansing beyond what is attainable using standard ETL and database matching functionality. The supreme scalability and fuzzy matching algorithms available through enterprise search technology allow users for the first time to create systems that maintain data consistency and cleanliness through familiar ETL GUI environments.

Once the data is cleansed through this technology, it can then be written back for use in database transactional systems and applications, improving their accuracy and efficiency.

Analyst Viewpoint

Organizations are increasingly deploying data profiling tools to attack the data quality problem at its source, in transactional systems. Data profiling is a sound first step in any data quality initiative and can provide greater confidence in data integrity before the information is channeled to a data cleansing application. In a data profiling initiative, IT professionals can conduct a fairly granular analysis of the content, quality, and structure of data in transactional systems (or in target data warehouses or business applications). Properly executed, data profiling can identify and correct systemic problems in source data, better prepare the information for such data cleansing techniques as algorithmic fuzzy matching, and help ensure the reliability of business data.

Request more information about FAST

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

Unification is the distinguishing characteristic of a truly mature BPM implementation. To varying degrees, many organizations have implemented BPM systems in what amounts to a tactical fashion, often in support of finance or sales functions. A unified BPM deployment will cover all aspects of operations, such as manufacturing, logistics, customer service, and marketing. Only a comprehensive BPM system fully enables an organization to look under rocks that may conceal critical weaknesses in operational performance. Most organizations will evolve only gradually toward unified BPM because the transition usually involves reengineering at an architectural level, but they stand to reap the benefits of improved visibility and the ability to execute on strategic objectives.

Request more information about Hyperion Solutions Corporation

Informatica Corporation

How do we manage the political minefields that commonly arise when establishing a data governance program?

Certain organizational principles are common across successful data governance programs:

- Strong sponsorship from a senior business executive.
- Rigor in defining roles and assigning specific responsibilities to individuals to enforce accountability.
- Equal business and IT involvement, which means that the business “owns” the data and takes the lead in defining policies, and IT partners with the business to help implement the supporting data integration technology.
- Establishment of an integration competency center (ICC) to promote reuse, share best practices, and establish common processes and standards for data integration.

Analyst Viewpoint

It's seldom easy to reduce political friction, whether it exists among business units or between business and IT. In the latter case, data governance organizations should consider borrowing from a BI systems management approach that puts a premium on individuals with both business acumen and technical skills. Cross-pollinating business and IT skills, so that IT personnel have a better appreciation of the business side and business users develop basic technical skills in such areas as data definitions, can help defuse political landmines and better align business and IT. An ICC, headed by a senior executive functioning as a data governance “czar,” offers an effective platform for the sharing of business and IT skills.

Request more information about Informatica Corporation

Mentor Graphics Replaces 1,200 Spreadsheets with a Single Web-Based Solution

Commentary by Jan-Willem Beldman, Team Lead for Analytical Applications and Data Quality, Mentor Graphics

Summary

Mentor Graphics has relied on Hyperion solutions since 1998, when they first chose Hyperion Essbase for global sales reporting. In the years that followed, the company has aggressively added to their solution with Essbase applications for executive reporting in 2001, budgeting and planning in 2002, and their most recent addition in 2005—a powerful system for managing worldwide human resources. Today, there are more than 300 users, and the company has a clear vision to make information even more broadly available in the coming years.

About Mentor Graphics

Mentor Graphics Corporation is a world leader in electronic hardware and software design solutions, providing products, consulting services, and award-winning support for the world's most successful electronics and semiconductor companies. Established in 1981, the \$700 million company employs approximately 3,950 people worldwide.

Managing the Most Important Asset

As a fast-moving technology leader, Mentor Graphics firmly believes that people are its greatest asset. With employees across 28 engineering sites and a global distribution channel with 48 offices around the world, the company must keep a constant pulse on global staffing for both the finance and human resource (HR) departments. As with its other data-intensive requirements, Mentor Graphics turned to Hyperion to quickly meet the needs of both. To best understand, analyze, and act upon global staffing needs, Mentor Graphics implemented a Hyperion Essbase multidimensional cube.

With Hyperion Essbase, almost immediately, the HR department realized the ability to make better HR allocation decisions. Some teams, for example, oper-

ate more as service organizations, with fluctuations occurring as project needs increase and decrease. Hyperion Essbase allows these groups to address the need to have 20 people work for three months on one project and then work for another group on a completely different project.

“By making more informed staffing decisions, we are growing top-line revenue by allocating people where the need is greatest, and managing the bottom line by making sure they are being deployed optimally at all times,” says Jan-Willem Beldman, team lead for analytical applications and data quality at Mentor Graphics.

Transforming the Budgeting and Planning Process

After its initial success with the Hyperion-based sales reporting system, in 2002, Mentor Graphics began an effort to replace the time-consuming planning and budgeting process—an annual event involving more than 1,200 spreadsheets for each worldwide cost center—with a bottom-up, Web-based solution using Essbase. Today, Mentor Graphics plans quarterly instead of annually, using a bottom-up planning process that ensures that data is collected accurately as each person moves through the process. With an automated workflow for each cost center, plans are approved hierarchically—first by division, then by region, and finally by the company controller. By the time the planning process is complete, there is a comprehensive model in place that includes details for each cost center and rolls all the way up to the top.

Ensuring Data Consistency and Version Control

Hyperion Master Data Management (MDM) is integrated with each of Mentor Graphics' Hyperion Essbase applications, allowing them to manage hierarchical master data across these applications, as



well as their transactional, reporting, and analytics systems.

“Today, we use Hyperion MDM as the central server to publish master data across heterogeneous systems from our multidimensional database, SAP, Calidus, and Oracle systems,” Beldman explains. “Hyperion MDM has enabled us to achieve consistency of our data and improve version control of various enterprise hierarchies.”

Constantly Improving Access to Information

Hyperion Essbase and Hyperion MDM provide the timely information, speed of analysis, and accuracy of data management that a geographically dispersed company like Mentor Graphics needs to remain competitive. To continue to improve the availability of information across the company, Mentor Graphics is currently exploring the use of dashboards as a way to bring timely, actionable data directly to the desktop.

“We have seen incredible results,” says Beldman. “Eight years ago, we had limited access to data. Today we have analytic power, worldwide access to information, true scalability, and unbreakable processes. Moreover, we have completely transformed the way we think about data. We are no longer dealing with a collection of abstract numbers, but we're building intelligence around every single aspect of our business. We make better, faster business decisions that directly benefit the bottom line and keep us competitive.” ■

BP Exploration and Production Migrates to a New Business Intelligence Environment

Commentary by Graham Yerbury, Head of MI Systems E&P Segment, BP

Situation

Providing accurate and timely forecasting and planning information is essential in the Exploration and Production (E&P) segment within BP. The market prices of oil and gas can and do fluctuate widely.

Overview

Industry

Manufacturing

Customer Profile

The Exploration and Production (E&P) segment of BP, one of the world's largest petroleum and petrochemicals companies, has offices in 26 countries.

Business Situation

BP E&P needed to upgrade its Group Financial Outlook forecasting and planning system to give its financial analysts more timely access to information through the corporate intranet.

Solution

Altius Consulting—BP E&P's consulting partner for this important project—recommended Microsoft SQL Server 2000 and Microsoft Analysis Services Business Intelligence platform, with arcplan dynaSight for Web-based reporting and analysis, and XLCubed for reporting and analysis via Microsoft Office Excel.

Benefits

- Saves \$100,000 in annual maintenance charges
- Improves controls and audit trail
- Enables faster, more flexible group forecasting and planning via the corporate Intranet
- Ease of integration with back-office applications

For More Information

www.microsoft.com/uk
www.bp.com/home.do
www.altius.co.uk
www.xlcubed.com

Production levels and resource requirements are under constant review. BP's financial analysts need to have access to information across a wide range of metrics, including financial, production, resources, and pricing. The information is updated regularly and quickly.

BP became acutely aware of the need to find a solution to replace its aging online analytical processing (OLAP) based Group Financial Outlook (GFO) solution, which was not as flexible and scalable as it needed to be. Having looked at competing options from Hyperion, Oracle, and Microsoft, BP turned to Altius Consulting, a Microsoft Certified Partner, for advice as to what solution they recommended given BP's need for rapid information turnaround and low cost of implementation and maintenance.

Graham Yerbury, head of MI Systems E&P Segment, BP, said: "We had an existing OLAP product, and in 2001 built a systems infrastructure around it for a local audience. Everyone who used it was in the same building, but within 12 months from implementation, we needed to extend the system to a broader audience and distribute it through the BP intranet. Unfortunately, the product we had did not give us a robust or scalable enterprise solution."

BP wanted to provide up-to-date data to its offices in Hong Kong, Houston, Buenos Aires, Aberdeen, and London to ensure that the analysts could access information at the same time. Using a file-based workaround in its legacy system, Yerbury felt that there was a high likelihood of information being out of sync.

Solution

Altius recommended that the new GFO solution be based on the Microsoft Business Intelligence platform using Microsoft SQL Server 2000 and Microsoft Analysis Services as the data storage and analysis engines. In addition, arcplan

dynaSight was recommended as the Web-based reporting and analysis tool. arcplan dynaSight is a Microsoft Gold Partner firm and leader in Web-based reporting and analysis tools. arcplan dynaSight provides users with a user-friendly way to navigate around their business information environment and gain a solid understanding of where their business is doing well and where it could do better. Features such as drill down and up, traffic lighting, and graphs make arcplan an intuitive work environment for BP's analysts.

For ad hoc reporting and analysis, Altius recommended XLCubed, a Microsoft Office Excel add-in that enables users to create reports on the fly utilizing the power and functionality of Excel and Analysis Services.

Altius has a strong relationship with Microsoft in the business intelligence marketplace. David Anderson, principal consultant with Altius, said: "Microsoft has changed the price profile for OLAP technologies for good. Companies can now gain access to OLAP and relational database technologies for a fraction of the cost of just a couple of years ago."

Microsoft Analysis Services is fully integrated with Microsoft SQL Server. Together these products have provided BP E&P with the necessary components for a fully integrated BI solution. arcplan dynaSight produces reports and graphics that export easily to Microsoft Office products, and the tool is extremely easy to use.

Graham Yerbury added: "dynaSight is a very robust enterprise application that fits in well with our intranet. We particularly like the tight integration with Microsoft Office for our monthly financial reports to the E&P management team. Previously, we had to print off tabs in 65 Excel workbooks. Now the information can be distributed as a Microsoft Office Word document through the e-mail system or read online." ■

Alaska Air Group Soars in Meeting the Needs of More than 17 Million Customers Annually

Commentary by Alaska Air Group

The Challenge

For 74 years, Alaska Air Group has been providing world-class domestic airline service. Understanding and responding to customer needs is key to its success in the increasingly competitive airline industry, where margins are small and customer loyalty is prized. The problem was not lack of data, but rather a wealth of it residing in disparate legacy systems. Alaska Air Group had data tracking customer behavior, but it was utilized in silos throughout the company. The existing systems lacked the ability to analyze data from the various sources to fully understand customer trends and purchasing behavior. Without a robust business intelligence solution, the customer tracking consisted primarily of examining lagging indicators, such as miles flown or dollars spent, rather than gaining an understanding of each customer's needs.

Prior to looking for a solution, Alaska Air Group looked at their customer data across the enterprise and at gaps in the process. One of those gaps was in accurate and timely information being made

available to executives for strategic planning purposes and to the tactical teams for executing to the strategic objectives. Alaska Air Group realized that what was needed was in-depth and timely analytical information that would allow a more strategic approach to addressing the needs of their customers.

Choosing a Solution

When Alaska Air Group defined the project, the focus was on understanding the customer's value chain, the information needed, identifying sources of data, and then finding a solution that would enable them to gain relevant, actionable insight from all these data sources—not just information that would be an aggregate, but something that would add value to their business process. Of the four solutions evaluated, Oracle BI Suite Enterprise Edition fit these requirements perfectly—it was the right solution, at the right price, with the right sophistication, yet very easy to use. Best of all, it only took about six weeks to implement and begin realizing this solution's benefits.

Improved Customer Loyalty

Loyalty is an indicator of how well a customer is being served, and includes measures like recent travel, frequency of travel, money spent, frequent flyer membership, and total mileage flown. In the past, it was difficult to get a clear picture of the factors that affect loyalty. With Oracle BI Suite, Alaska Air Group not only better understands the factors that affect loyalty, but also can get actionable, guided analytics to help improve it.

Relevant Customer Experience

Alaska Air Group continuously strives to provide a highly relevant experience for each customer. Providing a relevant experience to millions of flyers in numerous demographics is difficult. With Oracle BI Suite, the group can now combine



their customer interaction with the demographic data to provide the right promotion at the right time to the right person.

Improved Business Operations

Oracle BI Suite is enabling Alaska Air Group to make more strategic marketing campaigns possible. For example, they have started analyzing the reasons for seasonal “soft” markets and creating target offers to proactively offset them, as well as improving business processes by analyzing and understanding the correlation between business operations and customer initiatives.

A Solid ROI

With Oracle BI Suite, Alaska Air Group is able to gain a better understanding of the needs of its 17 million flyers. The goal was to integrate data across the enterprise, but what the solution is truly doing is driving organizational change from the bottom up. People are starting to realize that there are causes and effects between every element of a business, and this is helping in breaking down those silos within the organization. Thanks to a forward-thinking and progressive approach, Alaska Airlines is able to gain competitive advantage by driving actions, decisions, and business process improvements with deep customer insight. ■

Overview

Industry

Travel and Transportation

Oracle Product

Oracle Business Intelligence Suite Enterprise Edition

Key Benefits

- Drive business process improvement based on pervasive insight
- Better understanding of customer behavior and needs
- Drive strategic actions based on deep customer insight
- Scalable for users and data loads

Nationwide Insurance Transforms the Business with Innovative and Award-Winning Finance Data Management Initiative

Commentary by Lance Dacre, Director, Data Governance, Nationwide Insurance

The Challenge

Nationwide Insurance is one of the world's largest diversified insurance and financial services organizations. With hundreds of legacy applications scattered across dozens of lines of business, functional units, subsidiary companies, and geographies, Nationwide faced issues common to many other large enterprises—inconsistent data, poor visibility, and subpar data governance and reporting capabilities.

From the CFO's office to departmental analysts, Nationwide personnel struggled with contradictory data across multiple transactional systems for planning, budgeting, forecasting, risk management, and global consolidations that compromised their ability to make fact-based decisions. The company faced multiple "versions of the truth" as a result of system proliferation, including:

- 12 reporting systems
- 14 general ledgers
- 17 data repositories
- 20 charts of accounts
- 300,000 Excel spreadsheets

For a large, data-driven company, Nationwide's problems were not particularly unique. What was unique was Nationwide's far-reaching vision in addressing the challenge. It incorporated advanced capabilities to manage and ana-

lyze metadata—the proverbial “data about data”—as a key foundational element in a revamped finance data management platform that it called FOCUS.

By building FOCUS atop a metadata-driven data integration architecture, Nationwide foresaw opportunities for a single, enterprisewide system that consolidated metadata from multiple systems to give both business and IT personnel a powerful and incisive toolset to better understand and manage complex business dynamics. By using metadata, Nationwide recognized that it could:

- Build confidence in the data that drives the business by tracking data lineage, managing data governance, and ensuring data validation
- Improve business users' understanding of interrelated data across systems
- Improve data visibility, quality, and financial reporting
- Simplify change management and impact analysis for IT
- Slash the time and expense of manual data review and reconciliation

The Solution

Nationwide chose Informatica PowerCenter Advanced Edition (AE) because of past success with the data integration platform in several smaller deployments, its broad scalability, and its proven ease of use. Nationwide found PowerCenter AE's robust metadata management and analysis capabilities an essential ingredient for cross-enterprise data visibility.

PowerCenter provides the data integration backbone to source from 200 legacy applications—a diverse collection of homegrown systems, packaged software, and mainframe flat files—to an NCR Teradata-based financial data warehouse with some 450 gigabytes of data. More than 2,000 analysts across mul-



iple business units use Hyperion business intelligence tools and an Essbase OLAP engine for reporting and analysis against the warehouse, a data mart, and eight multidimensional OLAP cubes.

Metadata Management Benefits for Business and IT

A key feature of PowerCenter AE called Metadata Manager helps to ensure data consistency across the FOCUS ecosystem by giving business and IT users new capabilities to understand and adapt to data changes across systems. By ensuring the quality of business, technical, and operational metadata, Nationwide shares a common vocabulary so that data is created and understood based on standardized definitions.

Connectivity to metadata sources was established through prebuilt PowerCenter connectors, as well as custom metadata configurators. Built-in capabilities for authentication and authorization govern which users have access to what data, based on roles and responsibilities. Data lineage provides the ability to track who changed what data, where, and when, to help satisfy auditing requirements.

PowerCenter extracts from the metadata sources—amounting to about three million rows of data—every four hours to update a dashboard used by about 60 Nationwide business users to track field definitions and metrics in the information delivery layer. These self-service capabilities enable business users to substantially reduce the time required to validate, analyze, and report on financial processes. They also reduce the time IT spends handling business user inquiries.

Definition of Data Governance

The processes, policies, standards, organization, and technologies required to manage and ensure the availability, accessibility, quality, consistency, auditability, and security of data in a company or institution.

“Nationwide has achieved a single, authoritative source of enterprise finance data, resulting in more accurate and reliable forecasting, streamlined operations, and a greater ability for the finance department to serve as a true business partner and strategic advisor to groups across the company.”

—Lance Dacre, Nationwide Insurance

Separate weekly extracts consolidate FOCUS technical metadata into a single, searchable repository. With PowerCenter AE, about 40 Nationwide IT professionals have realized breakthrough capabilities for impact analysis and change management. With Metadata Manager, PowerCenter AE provides clear visibility into data relationships and the ripple effects of change, and helps identify redundancies and streamline data flows and models.

Importantly, PowerCenter AE provides visibility and reporting into Nationwide’s Kalido master data management system. It propagates master data across the ERP financial systems, data warehouse, and data mart, and provides a comparison of Kalido-generated master data dimensions and hierarchies across the target systems.

The first two FOCUS iterations went live in February 2006 to cover planning, budgeting, allocations, forecasting, risk management, regulatory reporting, and other functions across five subsidiary companies. Concurrent with the FOCUS rollout, Nationwide built out a financial data governance organization that included business and IT stakeholders, and implemented processes that defined ownership and responsibility for data and metadata across the company. The organization relies on PowerCenter AE to help establish, monitor, and enforce data governance policies.

The Results

Transform the Business with Timely, Reliable, and Accurate Information

With PowerCenter AE, Nationwide has improved its ability to track, analyze, and manage business dynamics across its highly complex data architecture. Metadata Manager provides access to a comprehensive metadata repository that enables highly granular analysis of lineage and change impact. That visibility is indispensable in satisfying auditability requirements, complying with Sarbanes-Oxley, and instilling confidence in the integrity of Nationwide’s financial information, while reducing by days the time required for manual review and rework.

Reduce Costs by Automating Manual IT and Business Processes

PowerCenter AE has eliminated costly and time-consuming chores for both business users and IT professionals. Prebuilt connectivity and ease of use has substantially reduced the time IT administrators spend maintaining disparate systems, reconciling discrepancies, and custom-coding new interfaces. In the business, Nationwide has cut the number of days required to close its books in half—from 15 days to 7.

“Future-Proof” Extensibility for Business Growth and Adaptation

Built on open standards and a flexible architecture, PowerCenter positions Nationwide to rapidly extend its FOCUS financial platform to meet such challenges as mergers and acquisitions and tightened integration with related customer and product data management systems. For instance, Nationwide is in the process of retiring a set of legacy UNIX scripts used for back-end ETL control processes and migrating those to PowerCenter AE.

Gain Insight and Competitive Advantage by Consolidating General Ledgers and Procurement

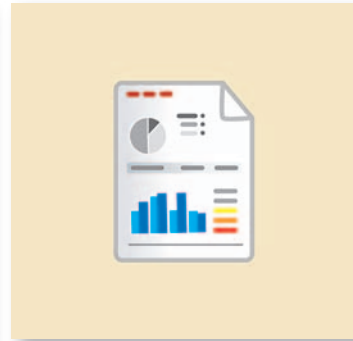
On the finance side, the PowerCenter-based FOCUS platform has enabled the consolidation of 14 general ledgers into a single PeopleSoft general ledger used by more than 150 people. PowerCenter’s flexibility has enabled Nationwide to add new interfaces that increase the number of sources and load frequency. On the spend side, FOCUS has enabled the consolidation of multiple procurement systems. Now, a single view of suppliers and a transparent purchasing process have positioned Nationwide to negotiate better deals with vendors.

TDWI 2006 Best Practices Award

For its innovative FOCUS finance data management initiative, Nationwide won a TDWI 2006 Best Practices Award in the master data management category. [See page 55.] “FOCUS has had a profound business impact across all aspects of Nationwide’s operations, literally reinventing the way finance services are performed and delivered to the enterprise,” said Lance Dacre, director of data governance for Nationwide. “Nationwide has achieved a single, authoritative source of enterprise finance data, resulting in more accurate and reliable forecasting, streamlined operations, and a greater ability for the finance department to serve as a true business partner and strategic advisor to groups across the company.” ■

Deploying Dashboards and Scorecards

By Wayne W. Eckerson,
Director of Research and Services, TDWI



Why Performance Dashboards?

Convergence of Two Disciplines.

Dashboards and scorecards represent the convergence of two distinct disciplines in desperate need of each other: performance management and business intelligence. Like long-lost siblings, these two disciplines have struggled on their own to deliver real business value and gain a permanent foothold within organizations. But together they offer an unbeatable combination whose whole is greater than the sum of the parts.

Performance Management. Performance management is the process of measuring progress toward achieving key goals and objectives in order to optimize individual, group, or organizational performance. Performance management encompasses strategy-setting, goal-setting, planning, budgeting, forecasting, and modeling techniques.

Business Intelligence. Business intelligence (BI), on the other hand, consists of the tools, technologies, and processes involved in turning data into information and information into knowledge to optimize decision making. BI encom-

passes data warehousing, data integration, reporting, analysis, and data mining technologies.

Together, these two disciplines provide a powerful new way to communicate strategy to all employees and monitor and analyze business activity designed to optimize performance. The result is a dashboard or scorecard—a new type of performance management system—that uses BI technologies to apply performance management techniques on an enterprise scale.

Three Major Benefits. When properly deployed, dashboard and scorecard systems offer three main benefits:

1. Communicate Strategy. They provide executives with a powerful means to communicate key strategies and objectives continuously by tailoring metrics to each employee based on his or her role and level in the organization. As an agent of organizational change, dashboards and scorecards enable executives to get the entire organization marching in a coordinated fashion toward the same destination.

2. Monitor and Adjust the Execution of Strategy.

Once goals are established, dashboards and scorecards let executives and managers monitor the execution of the strategy and plans on an hourly, daily, weekly, or monthly basis depending on requirements. These performance management systems enable executives and managers to work proactively and identify and address critical problems undermining progress before it's too late to fix them.

3. Deliver Insights and Information to All.

Dashboards and scorecards deliver critical information at a glance using graphical symbols, colors, and charts. The applications graphically highlight exception conditions and alerts and let users drill down into more detailed data to find the root cause of a problem. These tools conform to the way users work and don't force users to conform to the way BI tools work. This is a major reason dashboards and scorecards are so popular today.

Dashboards versus scorecards

	DASHBOARD	SCORECARD
Purpose	Measures performance	Charts progress
Users	Managers, staff	Executives, managers, staff
Updates	Real-time to right-time	Periodic snapshots
Data	Events	Summaries
Top-level Display	Charts and tables	Symbols and icons

Figure 1. Dashboards and scorecards are visual display mechanisms—the monitoring layer in a performance dashboard. A good performance dashboard should support both types of displays.

What Are Dashboards and Scorecards?

The many different types of dashboards and scorecards can each look and function slightly differently, and often go by different names depending on the organizations implementing them. To dispel the confusion, here is a definition:

Dashboards and scorecards are multilayered performance management systems, built on a business intelligence and data integration infrastructure, that enable organizations to measure, monitor, and manage business activity using both financial and non-financial measures.

Dashboards and scorecards provide more than just a screen populated with fancy performance graphics: they are full-fledged business information systems designed to help organizations achieve strategic objectives. They help measure the past, monitor the present, and forecast the future, allowing an organization to adjust its strategy and tactics in real time to optimize performance.

The Difference Between Dashboards and Scorecards

Some people use the terms dashboard and scorecard interchangeably, while others use the terms to refer to different types of analytic applications for measuring performance. In my definition, dashboards and scorecards are simply different types of *visual display*

mechanisms, within a performance management system, that convey critical performance information at a glance. In other words, they are the monitoring application, not the system itself!

A good performance management system should be able to deliver either a dashboard or scorecard interface, since both do the same thing: display the status and trends of key performance indicators. The primary difference is that dashboards tend to monitor the performance of operational processes whereas scorecards tend to chart the progress of tactical and strategic goals. Dashboards also tend to display charts and tables with conditional formatting, whereas scorecards use graphical symbols and icons to represent the status and trends of key metrics. (See Figure 1.)

Best Practices in Deploying Dashboards and Scorecards

Much of the advice in this section can be applied to all IT projects, but you will also find targeted tips and techniques for deploying dashboards and scorecards.

1. You Get What You Pay For. You can't deploy a dashboard or scorecard that delivers real and lasting business value without spending money—unless of course, your company has already invested in a BI and data management infrastructure that can deliver the right information to the right user at the right time in the right format, and has already modeled and loaded a

majority of data required in the dashboard or scorecard. Here, companies that have successfully deployed a data warehouse and BI tools have an edge. Those that haven't incur greater costs.

One IT manager at a large manufacturing and distribution company garnered significant attention internally when his team created an effective (but decidedly low tech) dashboard for an executive vice president to monitor customer service operations. Soon, other business managers approached the IT manager to convert their Excel and Access databases into dashboards.

“They were decidedly uninterested when we told them how much it would cost,” says the IT manager. “They don't understand the costs involved in cleaning, integrating, and modeling the data and building a bulletproof system that delivers sufficient right-time and detailed data so that they can make accurate and timely decisions.” The IT manager said it cost \$400,000 to build the dashboard and that it's been running for four years without any dedicated IT support.

Of course, you can deploy an inexpensive performance dashboard today, even without the requisite infrastructure. There are many commercial products that you can license for under \$50,000 that provide some data integration capabilities. However, you will eventually need to replace these products as word spreads about the success of your solution and you need to scale it up to support more users, more sources, more detailed data, and more frequent updates without compromising performance and response times.

2. Plan for the Long Haul. Always plan for success, because the alternative is disaster. Word about successful dashboard and scorecard solutions spreads like wildfire. If you've delivered a successful solution, you'll be bombarded with requests to deliver performance management systems for other departments and will need to rapidly expand the scope and scale of the existing system. Meanwhile, the number

of users may grow rapidly, placing undue burden on processing power, networks, and databases. If you are not careful, response times will plummet and your hard-won reputation will suffer irreparable damage.

Chris Gentry, director of business intelligence at CCC Information Services, offers this sensible advice based on years of delivering BI solutions: “Unless you prepare for 20 percent growth in users, 15 percent growth in queries, and four to five new data sources every year, you will not meet customer expectations. The initial footprint of your solution should be 15 percent more than your most optimistic forecast.”

3. Plan for Real Time.¹ The value of a dashboard or scorecard increases exponentially with the freshness of data. This is not to say that dashboards or scorecards that are updated monthly don’t deliver business value—they can! Many successful Balanced Scorecards, for example, are only updated monthly. However, augmenting a dashboard or scorecard with more timely data increases its value even more. A performance management system populated with more timely data lets executives and managers keep their fingers on the pulse of the organization in ways they never could before. They work much more proactively to optimize performance.

So, even if your business people don’t ask for more than daily updates, be prepared to deliver them. Build in hooks to messaging backbones, if they exist, to trickle feed data into your environment. Or select commercial dashboard solutions that support event-driven processing and

can prove their scalability across users, sources, and data volumes.

4. Develop on a Single Platform. Because “dashboards are in the air” as one BI manager said, it is very easy for managers to build or buy their own solutions independent of each other. The dashboard silos eventually compete with each other for resources and, more importantly, undermine an organization’s ability to get a single picture of what’s going on in the enterprise.

This is what happened at IBM several years ago. The company had many different dashboard solutions serving different IBM executives in various departments. Each ran on a different platform, contained overlapping data, and duplicated resources, according to Julia Kennedy, senior manager of enterprise BI and reporting in the CIO’s office of IBM. Eventually, IBM executives selected one dashboard among many to standardize on across the company. The winner, a solution called EDGE (Enabling Decisions for Global Execution), now supports 22,000 users throughout the enterprise.

To avoid the disruption of switching to a new platform, it’s best to develop all dashboards and scorecards on a single platform that leverages a unified data integration infrastructure. The dashboard platform should support team-based development with check in/out and version control with seamless support for the deployment of development, testing, and production environments. The data infrastructure ensures the delivery of trusted business data using data integration, data quality, and metadata management tools. These technologies

enable business users to examine the lineage of any data element before making critical decisions. Also, administrators can assess the impact of changes to source systems or applications on data integration workflows and reports.

While it’s possible for a single development team to stay ahead of business requests for new performance management systems in a small company, it’s likely that the organization may need to appoint a program office to manage deployment across multiple departments using parallel development teams. The program office needs to establish standards for usage, data, metrics, and functionality to avoid creating non-integrated silos.

5. Develop Effective Metrics. Again, metrics are ultimately the key to the success of any dashboard or scorecard. They are the linchpin between the business and technical architectures. There are many techniques for developing effective metrics, but these are the most important ones:

A. Get Buy-In. Metrics won’t propel the organization in the right direction unless people whose performance is being measured understand, accept, and endorse the metrics. The best way to gain this buy-in is to involve workers in the process of defining metrics, targets, and thresholds. Since workers are closer to the business processes, they will better understand whether metrics accurately capture the nuances of a process and whether targets are realistic or not. When you gather requirements from workers, get their feedback on proposed metrics and targets. You’ll design more accurate metrics that have a greater impact on the business and are less likely to be manipulated by individuals who now have a vested interest in making them work.

¹“Real time” is an ambiguous term that means different things to different people. Real time can refer to data that is refreshed instantaneously, every minute, every hour, every day, and so on. For this reason, many people prefer the term “right time” because it emphasizes that data only needs to be as fresh as the business process requires.

B. Simplify. Humans can only absorb so much information at once. To avoid cluttering a dashboard or scorecard, display only a handful of metrics (four to seven is a good number) on a single screen at a time. If you have more metrics than this—which is not unreasonable if you are measuring an end-to-end business activity—you should create hierarchies of metrics using folders, tabs, or drill downs to preserve the clarity and simplicity of the dashboard display.

C. Empower. What's the point of displaying metrics if you don't empower users to affect the outcomes? When designing metrics, you need to examine the context in which they are used and give workers license to make decisions—even unorthodox ones—that will improve performance. This requires reengineering business processes and delegating responsibility to people closest to the customer or the process being measured.

Concomitantly, organizations need to hold workers accountable for the outcome of the measures. If no one is accountable, the metrics won't have any impact on the organization. It's also better to hold an individual accountable even if a team manages the process or task being measured.

D. Avoid Perfectionism. It's difficult to design new metrics in a vacuum. While it's important to gather comprehensive requirements and map business processes and information flows, you will never really know how well a metric works in practice until you deploy it and see what happens. To avoid analysis paralysis, adhere to the 80 percent rule: develop the metric to a point where you are 80 percent confident that it will have the desired effect, then deploy it, track the results, and refine it as needed.

E. Monitor and Revise. All metrics have a natural lifecycle. Over time, metrics lose their impact as workers streamline processes to the point where additional gains are not worth the effort. In addition, the business usually changes, forcing organizations to add new metrics that cause workers to expend their energies elsewhere.

To ensure the effectiveness of the metrics and the performance management system as a whole, you need to monitor the usage of the metrics on a continuous basis. Dashboard and scorecard teams should aggressively prune underused metrics (after first consulting with the business, of course), and they should monitor the uptake of newly deployed metrics to quickly identify problems that users may be having with the metrics or views.

F. Incentives. It's very dangerous to attach incentives in the form of bonuses or compensation to metrics that have not been fully vetted and accepted by the organization. It's important to deploy metrics for a while in order to identify and close potential loopholes, change calculations to more accurately reflect reality, and provide sufficient training so users understand how to impact outcomes. Attaching incentives to metrics prematurely is a recipe for disaster and can catapult organizations into chaos.

G. Involve Technical People. One common mistake is to create metrics for which no data exists. To avoid this awkward situation, make sure you assign technical people to the team that gathers requirements and designs the metrics. During discussion, these people can evaluate the existence and condition of data needed to populate proposed metrics.

Conclusion

The future of dashboards and scorecards looks bright. Today, many organizations have deployed performance management systems, but the size and scope of a majority of implementations is still small. This will change as user organizations gain more experience with these systems and vendors begin to offer robust commercial dashboard and scorecard solutions.

Growth is bound to happen because dashboards and scorecards provide business users with all the information they need to make effective decisions and achieve strategic objectives without being overwhelmed. These systems deliver data on demand as users need it—from visual displays that let users monitor KPIs at a glance to dimensional and operational data for conducting analysis and taking action. ■

Wayne W. Eckerson is the director of research and services for TDWI. He has 17 years of industry experience and has covered data warehousing and business intelligence since 1995. Eckerson is the author of many in-depth reports, a columnist for several business and technology magazines, and a noted speaker and consultant. His book, Performance Dashboards: Measuring, Monitoring, and Managing Your Business, was published by John Wiley & Sons in 2005. He can be reached at weckerson@tdwi.org.

This article was excerpted from the full July 2006 report. TDWI appreciates the sponsorship of ADVIZOR Solutions, Business Objects, Celequest, Corda Technologies, Hyperion Solutions, Microsoft, MicroStrategy, SAP, Oracle, and Visual Mining.

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

The Five-Step Approach to More Valuable Enterprise Data

By Daniel Teachey, Director of Corporate Communications, DataFlux Corporation

Many organizations are implementing new strategies to get ahead of the competition. But they frequently ignore the quality of their underlying data, leading to poor decisions, insufficient customer service, and ultimately, lost revenue. Some of the reasons for the explosion of inconsistent, unusable data include human error, poorly structured data entry procedures, and a lack of data standards across business units.

Before you begin any new data-driven initiative—such as enterprise resource planning (ERP) or customer relationship management (CRM)—it's important to address data quality issues within your existing data sources and to prevent your data from becoming outdated.

The Five Components for Data Improvement

Technology currently exists that allows businesses to improve and consolidate corporate information on an ongoing basis. This proven five-step data quality process helps these businesses analyze, improve, and control their data across the enterprise:

1. Data Profiling. Prior to building a house, an architect creates a detailed blueprint of the structure. Unfortunately, most data quality projects don't take this same approach to building better data. Data profiling helps you determine the current state of your data and allows you to discover problems with the data—including content, structure, and relationship problems across columns and tables. By understanding the strengths and weaknesses of your data, profiling gives you a starting point for making substantial improvements.

2. Data Quality. If data profiling provides a blueprint of data problems, the data quality phase begins the process of building better data. With data quality, you can correct errors, standardize information across tables, and validate

information that is inconsistent and inaccurate. Common data quality tactics include data standardization, pattern standardization, and address verification.

3. Data Integration.

Data about the same item often exists in multiple databases. This applies to virtually any type of data (e.g., customer data or product data). Once you have standardized and corrected data quality issues, you can more easily integrate this data across different sources. Data integration technology helps you intelligently integrate customer, product, and other kinds of business data across multiple sources.

4. Data Enrichment. Data enrichment builds on data quality and data integration by enhancing it with additional information and incorporating additional external data. With data enrichment technology, customer data can be combined and segmented more effectively to identify specific opportunities. Product, inventory, or financial data can be enhanced by adding additional bits of information from external sources. With data enrichment, you can create more valuable corporate information.

5. Data Monitoring. Organizations often believe their data quality problems are solved once they have improved existing data sources. However, maintaining high-quality data requires ongoing vigilance. Data monitoring allows you to create and enforce customizable, automated business rules to check the level of data quality over time. These rules, stored in a business rules repository and replicated across

the enterprise via Web services, allow you to enforce data governance rules and customize rules to be consistent with your operational processes.

Five Steps. One Platform.

The quality of any business analysis is only as good as the data at its foundation. Without data that is consistent, accurate, and reliable across the enterprise, an organization can easily reach misleading and potentially harmful conclusions.

With breakthroughs in technology, what was once an overwhelming endeavor can now be a streamlined data quality management process that is much less time-consuming, difficult, and costly. When implementing a data quality improvement program, look for a technology that can accomplish the five phases described here—all from the same interface. With a comprehensive platform of data improvement technology, an organization can elevate the role of data quality to an ongoing corporate activity, enabling better business decisions and a more competitive enterprise. ■



High-Speed Backup and Recovery of Large-Scale Data Warehouses

By Jesse Fountain, Director of Product Management, DATAlegro, Inc.

Overview

Backing up a large data warehouse is a tedious process that can disrupt data warehouse productivity for hours, if not days. Usually the data warehouse must be taken offline for this process, or worse, is run in the background, making the backup process take even longer. If data volumes are large, it can also be extremely costly, making it unfeasible for many companies. Moreover, new legal requirements such as compliance and other legislation, as well as growing business requirements, are forcing companies to find better options for quickly, efficiently, and safely backing up their data warehouse data.

DATAlegro B25 is a cost-effective solution to back up large data warehouses. The DATAlegro B25 has a 25-terabyte capacity, and multiple B25 units can be coupled together so there is room for even the largest data volumes. The DATAlegro B25 can back up data from a DATAlegro data warehouse appliance as fast as one terabyte every 12 minutes (based on “P” Series appliance) ... and restore it just as fast.

A Better Way of Backing Up Data

Many companies are forced to back up their data warehouse table by table. This is due to the difficulty of finding storage space for backup files and the time required to run the backups. The DATAlegro B25 takes a “snapshot” of an entire data warehouse image and compresses it quickly and efficiently into a single location without impacting the space available for the data warehouse. Better yet, full backups can be completed within relatively small and available batch windows.

Adding to the speed and efficiency of the DATAlegro B25 is the “differential engine,” which comes into play for subsequent backups. Instead of backing up the entire appliance each time, the differential engine looks only for those tables that

New legal requirements such as compliance and other legislation, as well as growing business requirements, are forcing companies to find better options for quickly, efficiently, and safely backing up their data warehouse data.

have been added or changed and adds them to the backup set.

The DATAlegro B25 can safely store backup files for as long as needed in a RAID 5 or 6 protected environment. In addition, backup files can be copied from the DATAlegro B25 to tape without impacting the performance of the data warehouse. Restoration of data can be at a table level, node level, or database level.

A Staging Area for Exporting and Importing Data

In addition to backing up data, the DATAlegro B25 can be used as a staging area for exporting and importing data to and from the data warehouse. Technically speaking, data from ETL tools such as Informatica can be “landed” onto the B25 and then, within the high-speed Infiniband network, data can be loaded or “upserted” rapidly into the data warehouse on the DATAlegro P or C series appliance. Finally, data can be exported from the DATAlegro appliance to external environments such as SAS as well as other large-scale data marts.

Repartitioning and Expansion

Growing your DATAlegro appliance environment is fast and easy with a DATAlegro B25. When growing data volumes result in the need to add another DATAlegro data warehouse appliance or additional nodes, the DATAlegro B25 can be used as the destination for a complete export of the data warehouse. Once the export is complete, the data is deleted from the existing appliance. Next, additional capacity is added to the appliance. Finally, the exported data is repartitioned and loaded across the newly expanded appliance.

Using DATAlegro B25 for expanding an appliance or repartitioning data provides an efficient and complete solution for maintaining and growing a large-scale data warehouse environment. ■

Who, What, When, and WHERE

Greater ROI from DW and BI Assets

By Steve Trammell, Corporate Alliances, ESRI

Nearly all organizations, both public and private, have a valuable but underutilized asset in their data warehouses. Furthermore, most organizations already have many of the tools necessary to exploit this asset and realize more ROI from those same data warehouses. This asset is location: the WHERE in the title of this article. While valuable analysis has been done on who, what, and when, the analysis of “where” has mostly been overlooked or confined to departmentally isolated groups.

The analysis of “where” is accomplished with geographic information system (GIS) applications. Over the years, organizations have discovered many valuable uses for GIS, such as sales territory alignment, risk analysis, target marketing, supply chain management, warranty analysis, fraud detection, and route optimization. Most organizations would benefit from many of these proven applications, but private sector organizations rarely implement more than one or two of them.

You Already Have the Data and Skills

Geographic analysis is fueled by location data, which is already present (in the form of addresses) in approximately 80 percent of operational data aggregated in data warehouses. This figure is even higher with business data. In addition, we have spoken over the past two years with a number of companies using BI, and have discovered that a surprisingly high number already have GIS implementations existing somewhere in their organizations. Not surprisingly, it was obvious that the BI users we spoke with had no knowledge of their organizations’ existing GIS usage. The reason is simple: GIS had been deployed in these organizations as an important tactical solution to a specific departmental problem. The data warehouse and BI users were working at

different levels within the organization: the operational data was being collected at the departmental level, so the GIS users of the source data had no impact on the DW and BI projects.

An example of this would be a real estate department that uses GIS to analyze potential sites for new facilities. The result of each analysis moves up the organization as a GIS-generated report, often with maps, until a decision is made. But rarely is anyone outside of the process exposed to the underlying technology, so other departments are not aware of potentially valuable in-house skills and applications. The lesson here is that practically every organization has the data to fuel geographic analysis, and most organizations have personnel that already understand and use the technology.

GIS and BI Integration Right Out of the Box

Today’s data warehouses and BI applications make it possible for many departments within an organization to benefit from geographic analysis. The data warehouse aggregates data from many operational units, so the GIS needs to connect only to a single source. BI provides a user-friendly analysis and reporting platform to simplify the actual application



of geographic analysis and map-enhanced reporting. GIS applications are now server-based, so all three technologies (data warehousing, BI, and GIS) can work together in an integrated IT environment. Realizing the value of GIS for enhanced reporting and forward-looking analytics, the major BI providers and their partners have built the linkages to GIS into their latest releases. In many cases, creating an interactive map-based report showing status, performance, or simple locations, is no more difficult than setting up a new report in your BI system. Now that you have the data analysis and distribution platform, it’s time to start exploiting the real value of WHERE. ■

Practically every organization has the data to fuel geographic analysis, and most organizations have personnel that already understand and use the technology.

Process-centric BI: Improving Business Processes with Comprehensive SOA

By Peter Caron, Senior Director, Global Marketing, arcplan, Inc.

The promise of a service-oriented architecture (SOA) is more than using a Web service to access a data source. The real value of SOA is in the improvement of business processes and the agility with which an organization can respond to change.

One opportunity is to embed business intelligence (BI) within processes for the first time. BI grew up consolidating data from operational systems, and making analysis available in pockets of the organization. The problem is that BI became data-heavy and inflexible, creating analytical stores that are unconnected and unable to interact with an entire business process. With Web services in an SOA environment, these walls are coming down.

Let's look at an example of this process-centric BI in improving cash flow and sales forecasting. A company has implemented the following independent BI applications:

- Cash flow analysis and forecasting
- Accounts receivable (AR) analysis
- Sales analysis and forecasting

Each application uses different data sources: the cash flow and AR applications primarily analyze the company's ERP system, while the sales analysis application uses data from the CRM system and a sales data warehouse. In addition, an external news analysis application, which is accessible as a Web service, is used by sales staff to research news on customers and prospects.

A process-centric solution connects these disparate sources. For example, a dashboard application is deployed against these four applications, making the applications and their associated metrics accessible via one interface. This connection means that alerts can be tied to an entire business process instead of just the data in one system.

The CFO, for example, may set up the following business monitoring rule:

1. Raise an exception alert if cash flow is lagging behind forecast by more than 10 percent.
2. If Alert 1 is triggered, drill into accounts receivable and check for forecast deviations.
3. If a significant deviation is found in accounts receivable:
 - Drill into sales and check for forecast deviations
 - Trigger a news analysis on the relevant customer or industries
 - Notify the relevant account executives on external trends

One of the benefits of this process-centric BI approach is faster time to action. The old approach raised independent alerts with no context, requiring expensive meetings and manual research to discover the complete issue. With process-centric BI, a financial analyst would receive an e-mail alert noting that:

1. A cash flow exception has been raised...
2. due to overdue payments...
3. and lower sales revenue within a specific customer segment...



4. potentially caused by the a steep rise in commodity prices, which had a strong effect on the particular customer segment (news analysis).

Possible action includes improving cash flow analysis by introducing sales or commodity pricing variables in the forecast. And if the dashboard has other process-centric capabilities like write-back to the source systems, then this reaction is automated, improving the efficiency of the process even more.

With comprehensive SOA integration, your BI environment adheres to your business processes rather than creating analytical silos. ■

The real value of SOA is in the improvement of business processes and the agility with which an organization can respond to change.

Key Requirements for Enterprise Business Intelligence

By Mark LaRow, Vice President, Products, MicroStrategy

In the past, companies deployed BI applications as departmental solutions, creating islands of BI for a distinct user population and database. Early success with departmental-level BI has led to greater demand to expand BI across the organization. As enterprises open up their ever-growing data assets to everyone in the company, they are now faced with the challenge of doing so in a cost-effective and scalable way.

The solution to this challenge is enterprise BI standardization, where one BI platform delivers a single version of the truth through a unified user interface to all people across the company. There are five key requirements for enterprise BI:

1 Support for Five Styles of BI

As a company moves into enterprise BI, it must have a technology that can seamlessly deliver all five styles of BI with a single unified architecture and metadata. The five styles of BI are:

1. Scorecards and dashboards
2. Reporting
3. OLAP analysis
4. Advanced and predictive analysis
5. Alerts and notification

Equally important is the need for a single technology to provide all five styles as plug-and-play components on a single unified backplane and through a unified user interface. The unified backplane ensures that all styles of BI share common services and automatically build upon one another. The unified user interface ensures a common experience for all users, allowing them to seamlessly move between styles of BI without cumbersome application switching.

2 High User Scalability

An industrial-strength BI architecture must support thousands of users, interactively accessing the same reports and scorecards. While many BI vendors cite large user populations, they do not

mention that these high numbers are aggregations of many small, isolated BI applications, or that they require massive server farms.

To achieve the highest user scale, companies need a BI technology capable of reaching all business users through the interface of their choice; a zero-footprint Web interface that is instantly deployable to geographically distributed users; and a dynamic caching architecture. A dynamic, multi-level caching architecture provides increased performance as more people use the system, making each user's experience faster and more interactive, while imposing the absolute minimum load on database resources.

3 High Data Scalability

Companies are faced with an explosion in the volume and scope of data they collect and generate. The myriad of systems, including ERP, CRM, Web sites, sales force automation, SEC compliance, and supply chain management, contributes to this massive data influx. Enterprise BI must be able to access all these databases to be truly effective.

A ROLAP (relational OLAP) architecture is necessary to access all data to the full depth and breadth of the largest databases. While some reporting-specific BI technologies can access very large databases, they sacrifice user interactivity. Reporting tools can access large databases, but only with relatively simple queries and without any real user interactivity. On the other hand, traditional OLAP tools can provide a high degree of user interactivity and exploration, but only against comparatively small amounts of data, stored in cubes, that is extracted from the relational databases. Relational OLAP architecture provides access to the largest databases with the highest interactive performance.

4 User Self-Service

As the population of users and amount of data in the enterprise grows, user self-service becomes absolutely critical. Enterprise

BI requires levels of self-service that go beyond ad hoc report design. To achieve broad-based self-service, companies need a BI technology that offers "what you see is what you get" (WYSIWYG) design, eliminating user training in report design. Even more effective for user self-service is a BI technology that can drill anywhere, allowing users to seamlessly drill through any combination of data in the data warehouse.

5 Automated Maintainability

Finally, cost-effective enterprise BI requires the ability to update all reports automatically, reflecting continual changes to underlying business definitions, business structure, and database structures. With a small BI application, the effort required to maintain reports manually is modest. However, with enterprise BI, report maintenance becomes the dominant issue in total cost of ownership, and automated instantaneous updating becomes a critical requirement.

Only BI technologies with a dynamic metadata architecture can ensure that all reports are updated instantly and automatically, with a single version of the truth, at any scale. Reports are immediately brought up to date whenever any underlying object changes, requiring fewer IT personnel to deliver the maximum amount of BI.

Enterprise BI has the power to deliver timely business information to everyone in an organization. Given the changing landscape of reporting, analysis, and monitoring requirements, traditional reporting tools and non-integrated BI products are reaching their limitations. As companies consider the benefits of enterprise BI, they must carefully choose a technology that can meet the most demanding requirements and enable thousands of business users to simultaneously access terabytes of data with mission-critical reliability. ■

There's More to Business Intelligence Than Just Software

Without the input of business users, your BI solution will fail.

By Gabhan Berry, CTO, XLCubed, Ltd.

Business intelligence is all over the news these days. It seems that not a week goes by without a business intelligence-related headline being splashed across the industry press. As the hype grows, the stakes rise, and the anticipation for businesses to get “real intelligence” from their solutions mounts. Companies buy the best software they can afford, build it as quickly as they can, and expect the insights and intelligence to jump out onto their screens... and why not? Isn't that what every vendor promises? Business intelligence should deliver business insights, not data.

But there is more to good business intelligence than good software.

There is more to delivering a business intelligence solution than having IT departments build a few analysis cubes, purchase a reporting tool, and hand it over to end users. Designing and building a business intelligence solution is not overly complicated, but it does require some up-front thought and planning. With more and more business intelligence vendors proclaiming simplicity and ease of use, fewer and fewer business intelligence solutions seem to deliver actionable insights to the business.

Time and again, I have seen problematic business intelligence solutions. Sometimes the problems are related to failing software, but more often the cause is human error: human error in the form of poor business modeling.

In order for a business intelligence solution to discover insights and allow the business to act on its data intelligently, the solution has to model the business problem accurately. This seemingly obvious statement is all too often overlooked or not even considered. How can a model that does not reflect the real world be of any use to people in the real world? How

can a model that does not understand the problem be used to solve the problem?

Before designing and building a business intelligence solution, the business needs to take time to understand the exact problem they are setting out to solve. Next, they need to get the right people around a table to provide an accurate model of the business, and it is here that the biggest mistake is made.

The right people for modeling the business are not IT staff; they are the business analysts. Most IT staff are too far removed from the day-to-day running of the business to be able to build an accurate model of the business.

Business analysts need to be involved in a business intelligence solution. It is their job to use the solution to enable the business to act intelligently on its data; they are the users of the system. Airplanes are not designed without pilot involvement; racing cars are not designed without driver involvement; so why would you build a business model without the involvement of a business analyst?

And yet so many do.

A failing business intelligence project is one that segments the IT staff



from the business staff. Different groups of people with different objectives are rarely able to come together as one. The IT department may build “technically” good multidimensional analysis cubes, but they are not useful if they do not reflect the business accurately.

Business intelligence is about uncovering insights. Important bits of information that impact your business are buried within the data. Data is critical in business intelligence, but data alone is not enough. You need a good model, and that model is going to come from the business analysts. Once you have a proper model, then you can begin to design, build, and deliver a business intelligence solution that will deliver information and insights to your desktop! ■

In order for a business intelligence solution to discover insights and allow the business to act on its data intelligently, the solution has to model the business problem accurately.

Company Directory Information and Customer Records: The Schibsted Case

By Per Anders Waaler, Senior Software Architect,
and Davor Sutija, Vice President, Strategic Market Development, FAST

Introduction

All companies of a certain size or complexity face the same challenge: Vital information about the business is stored in multiple locations and in different formats. This goes for customer data as well as product data. Getting that single version of the truth about particular customer earnings or product groups becomes quite an effort. Thus enters the data warehouse.

Data warehouses, however, do not solve one challenge: that of merging data from disparate sources without common keys. Customer data is typically stored in several systems: CRM, ERP, call centers, and so on. For a company that has recently undergone a merger or acquisition, the challenge only increases.

Schibsted Search and Sesam.no

Imagine the challenges involved in being an Internet directory provider. In Norway, the base for Schibsted, a large publishing company, there are more than 20 different providers of subscription services. All telecommunications operators are obligated by law in this particular market to deliver at their own cost a list of subscribers to third parties that request such information. This legal requirement,

however, does not mention nor guarantee data quality. As a publishing company, all such telecom record sources for Schibsted were external. Because of this, Sesam.no, Schibsted's online local search and yellow pages/white pages service, probably faced even more challenges than most in merging their databases.

Schibsted Search's goal was to present on the Web (at Sesam.no) all listings for a given company or person in a single GUI block, combining and cleansing multiple records. But even when the underlying telecom providers had the correct names for their subscribers, abbreviations and contractions abounded. For example, our own company, officially registered as Fast Search & Transfer, does business under the trade name FAST. Similar inconsistencies are especially prevalent in restaurant, retail, and other companies where a public trademark is not the name under which the company was registered.

By using multiple secondary and tertiary data fields, such as addresses and phone numbers, the link between the trade name and the corporate identity can be established. In addition, spelling errors and variations in word order that occur during data entry in call centers or online can be corrected automatically before publication using analogous tools.

The Solution

Schibsted had already standardized on FAST as their search platform on the end user side of the Web, offloading database technologies that could not cope with the increasing traffic and ad hoc query load that an online enterprise generates, both in the number of users and the query complexity.

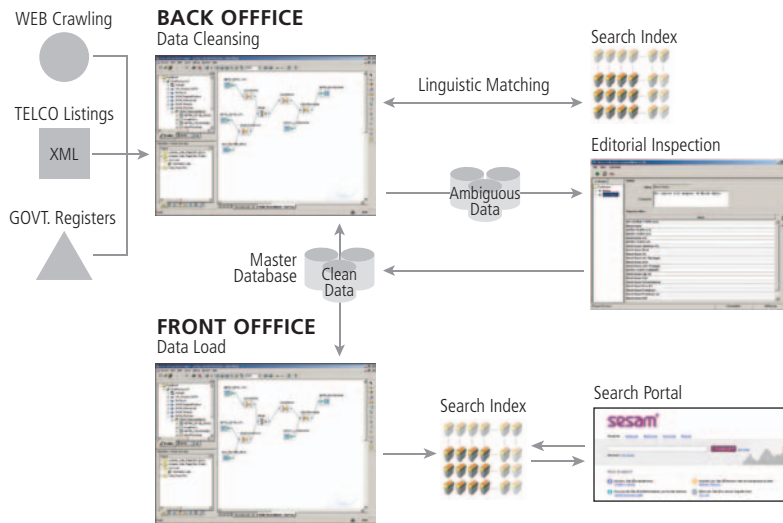
When we told Schibsted that they could also use the same technology for data cleansing, they were intrigued. As part of the FAST Data Cleansing Solution, adapters expose the FAST search engine as a database within leading ETL tools such as BusinessObjects Data Integrator and Informatica (when the complexity of the source data requires such a tool). The solution also includes embedded tools that can access popular RDBMS systems directly. In addition, FAST Data Cleansing can be called through custom APIs from any Java-enabled tool (TIBCO, webMethods, MQSeries, Oracle, and so on).

With this solution in place, we populated the FAST index from an arbitrarily large set of data sources, both structured and unstructured. Schibsted was interested in matching relevant directory listings to news stories and products using FAST's unique relevancy ranking tools.

The steps in the cleansing operation are straightforward: a data source is first indexed in FAST, and then additional records/sources are matched one-to-many using lookups in the FAST index. This process not only executes strict matches obtained by running queries against a database, but more broadly, uses various fuzzy matching algorithms. This approach, which reduces words to their core and then matches all forms of that core word, delivers comparative results ranked according to relevance and user-defined rule sets. All of this can be done easily within an ETL tool framework, using the adapter or specialized cleansing GUIs.

“The answer to our data quality issues, not only in data warehousing, but in all other aspects, is to build solutions using new tools such as the FAST Data Cleansing Solution to solve that issue. Regular database and programming logic could not do it for us.”

—Mikal Rohde, Schibsted Search



Data Cleansing Operation

- Read sources
- Track source changes
- Cleanse new/changed data against master
- Store result in database
- Allow for manual inspection GUI for unclear matches
- Update master and searchable apps with cleansed info

Different Scenarios

There are two basic scenarios that organizations face when attempting to merge data from multiple sources. In the first scenario, there is a master to cleanse against. This is easiest, but not the most common. In this scenario, we populate the FAST index with the master, and then cleanse other sources against that master, expecting to get a clear candidate for matching.

Scenario two is more typical. Here, there is no clear master, but instead a group of sources (such as a set of customer databases) that you want to cleanse against each other. The purpose is to find the same entity across disparate sources. In this scenario, we populate the master with one of the databases (chosen arbitrarily) and then cleanse that same set against itself to eliminate duplicates. Once any duplicates are removed, the cleansed set of database records acts as a master for further cleansing against the other repositories.

Cleansing Company Information

A typical scenario at Sesam.no was the cleansing of company information. This was accomplished using scenario one as described above. Since Norway provides a full listing and updates of all registered companies, this list can be used as the master to cleanse against. In the Oslo region alone, a metropolitan area of one million people, there are some 60,000 registered companies. Cleansing

in most cases is done on a combination of company name, street, zip code, and municipality.

Take a look at the sample case described in the sidebar below. With a potential of 60,000 companies, given a misspelled company name and the city, we are able to find the correct entity. The solution handles other typical quality issues equally well, such as words switching place (first/last name versus last/first name), synonyms, and lemmatization (recognition and matching of different grammatical forms).

When Schibsted and Sesam.no were awarded the Norwegian Data Warehouse

Award for 2006, Mikal Rohde, CEO of Schibsted Search, said: “The answer to our data quality issues, not only in data warehousing, but in all other aspects, is to build solutions using new tools such as the FAST Data Cleansing Solution to solve that issue. Regular database and programming logic could not do it for us.”

The jury of the Data Warehouse Conference, hosted by Confex Norway, stated in their evaluation that the differentiating factor for the winning entry was their solution to the problem of data quality, which allowed their data warehouse to drive their entire online business. ■

Sample case

Company name: beWise as (*this is actually a misspelled version of bWise as*)

Street: Not provided

Zip: Not provided

City: Oslo

Having populated the master index with all the 60,000 companies in Oslo, the (partial) result we get back as a candidate list using the fuzzy matching is:

RANK	Company	Street	Zip	City	Cleansed
1	BWISE AS	RÅDHUSGATA 9	151	OSLO	1
2	STREETWISE AS	LØKKEVEIEN 5	253	OSLO	0
3	MACWISE AS	ULLERNCHAUSSEEN 119	284	OSLO	0

The cleansed flag is set based on thresholds given to the fuzzy lookup function.

How to Launch a Data Governance Program: Practical Guidelines for Technology Implementers

By Judy Ko, Director, Enterprise Marketing, Informatica Corporation

Organizations are recognizing the value of data as a strategic business asset—one to be carefully developed and managed. However, many have not yet evolved their processes, policies, and infrastructure to cultivate their data as an asset. As a result, data governance is an emerging discipline that many organizations are only beginning to adopt.

Data governance is defined as the processes, policies, standards, organization, and technologies required to manage and ensure the availability, accessibility, quality, consistency, auditability, and security of data in a company or institution.

To derive maximum value from your data, you need a strong data governance program that helps you develop and manage your data as a strategic business asset. What can your IT organization do to launch a data governance program? The key is to take an incremental approach, ensure buy-in across both the business and IT, and build the right foundation. Here are four practical things to do to get started now.

1 Tie the Pilot to a Key Business Goal

When selecting a pilot for data governance, identify a project that will have a significant impact on the business. Data governance requires strong involvement from the business; tying the pilot to the

success of a business executive sponsor increases the visibility of the effort and helps attract resources and budget. And after a successful pilot, having documented business value facilitates efforts to expand the data governance program across the organization. Common projects that can deliver measurable business impact include:

- M&A consolidation
- Regulatory compliance
- Master data management

2 Make Pilots a Spreadsheet-Free Zone

When working on pilot projects, avoid the temptation to rely on the siloed tools your data stewards, architects, and analysts may use simply because they are familiar, such as Excel spreadsheets or standalone metadata modeling or data cleansing tools. The goal of the pilot project shouldn't be limited to developing the business standards and processes around data governance. The pilot should also be used to begin establishing the robust data integration infrastructure the organization needs to scale out the data governance program over time. Selecting a unified data integration platform that can help enforce policies and automate processes helps lower the cost and risk of ongoing data governance efforts.

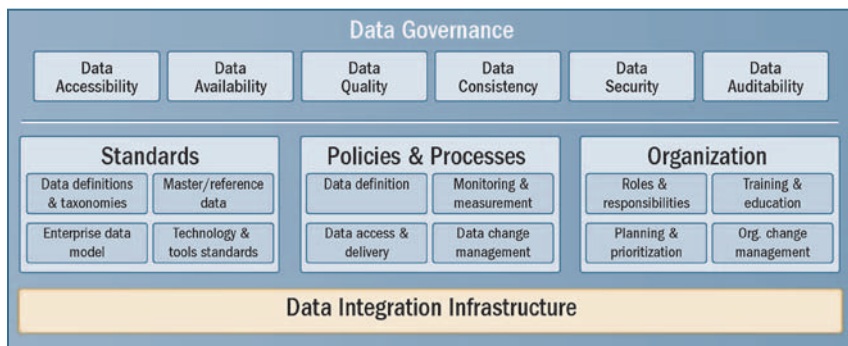
3 Talk to the Enterprise Architects

Make sure you sit down with the architects who own the overall architectural road map for the enterprise, particularly if these architects are not already intimately involved with the data governance program. It's important to understand their current architectural blueprint and future road map, and how they believe data governance plays into them. Be sure to talk about data services and how they fit into the larger architecture to ensure that the enterprise architecture is designed to support the goals of data governance.

4 Evaluate Technology Based on the Right Criteria

Developing the right technology infrastructure is critical to your ability to automate, manage, and scale out your data governance program. You need to identify a robust data integration technology infrastructure that can support your program's processes, policies, standards, organization, and technologies. When you evaluate a data integration technology platform, be sure to examine its ability to ensure that:

1. All enterprise data can be accessed, regardless of its source or structure.
2. Data is available to users and applications—when, where, and how it is needed.
3. Data is accurate and valid.
4. The value, structure, and meaning of data are consistent and reconciled across systems, processes, and organizations.
5. There is an audit trail on the data and internal controls have been appropriately implemented.
6. Data can be accessed securely. ■



Without the right data integration technology infrastructure, it's nearly impossible to ensure that data is accessible, available, of high quality, consistent, auditable, and secure across the enterprise.

It's Not About the Report—It's About What's In It!

By Frank Buytendijk, Vice President of Corporate Strategy,
and Howard Dresner, Chief Strategy Officer, Hyperion Solutions Corporation

Over the last few years, the business intelligence (BI) market has been amazing. The quest by business for better insight has been addressed by a singular focus on technology issues. One holy grail after the next has been introduced as the answer to achieving real insight. It started with “BI for the masses,” with the misguided notion of ad hoc query capabilities for every conceivable user in the company. Then, production reporting became popular, promising to provide incredible scalability—and again addressing everyone in the organization.

The latest solution is search-style access to BI. Asking a typical data warehouse for a report on “Revenue for Product Group X” will yield endless results, with many indistinguishable definitions of revenue and many different hierarchies that potentially make up Product Group X. Unfortunately, search today completely lacks the requisite intelligence. Does the user mean net sales revenue? Should the report include currency translation effects, treasury results, and other sources of income? Confusion is guaranteed. Is this progress?

New Reports—Same Old Data

A large percentage of business data is just not fit to be consumed by most business users, and significant enrichment is required to transform raw data into information. A business context is needed for the data to become meaningful. And this business context is provided by master data and metadata. Master data—which contains the approved lists for basic business entities such as all product codes, customer codes, geographies, timetables, and so on—needs to be standardized and aligned to be able to compare information on multiple levels of detail. Metadata—which contains the definitions of business terms such as revenue, market share, and number of items produced—is necessary

to make sure users don't just see what they ask for, but what they should see.

The Two Dimensions of Strategic BI

The outlook is not all bad, however. There is a significant advantage to being able to easily distribute relevant business information to a large number of people. But the consumption of management information by all user constituents is at best half the story. Strategic BI can only be achieved by addressing two dimensions: the strategic *production* of BI next to the strategic *consumption* of BI.

Thus far, the BI market has mostly been associated with the strategic consumption of data, namely, integrated reporting and query tools available for large numbers of users. For the most part, BI solution providers use the same approach—with a metadata layer that shields users from the database structure and a visual approach to building reports or queries. The truth is that for the consumption of data, there is next to no competitive differentiation between the different vendors anymore. The query and reporting capabilities of all BI vendors are largely the same.

The Engine of BI

It is the strategic production of BI that makes the difference. This is where analytics and business performance management (BPM) come in. These elements essentially comprise the engine of strategic BI. This engine is designed to create



actionable insight through an understanding of business scenarios and management processes.

Convergence of BI and BPM

The market trend is clear: BPM and BI, previously separate markets, are converging into a single entity at an amazing speed. Why? To enable both the strategic production and consumption of BI. In fact, by 2008, the market will not make a distinction between the two markets. Many of the established vendors will take a centralized approach, creating large product suites with comprehensive functionality. These suites, however, are heavily controlled by IT and seriously lack the flexibility for users to create all business scenarios.

We believe Hyperion is best positioned to benefit from the convergence trend. Hyperion's approach has been to go beyond the integration of either BPM applications or BI technologies. With Hyperion System 9, Hyperion combines BI—including all of the underlying analytics—with BPM applications. With the Foundation, System 9 offers the IT department everything it needs to manage the BI and BPM platform in a centralized manner. ■

Strategic BI can only be achieved by addressing two dimensions: the strategic *production* of BI next to the strategic *consumption* of BI.

Intelligent Data Warehousing with ETPL

By Keith Laker, Oracle Warehouse Builder Product Management, Oracle

Every organization is struggling to ensure the provision of intelligent data warehouse (*i*DW) environments that contain as much embedded value as possible. However, any embedded value is usually constrained by a number of typical data-related issues:

1. Missing or incorrect data
2. Lack of relevant business attributes
3. Lack of focus on most relevant attributes, hierarchies, and levels
4. Ability to located critical data in a timely fashion

Performing simple extract, transform, and load (ETL) operations as part of the data warehouse process is no longer sufficient.

Adding Value To The ETL Process

It is possible to add significant value to the data warehouse environment by incorporating predictive analytics directly within the normal data loading process. The whole ETL process simply expands to include a new, and critically important, stage—“P,” or *predict*.

By incorporating a predict stage as part of the upstream processing, all business users can benefit from the results, no matter which reporting tool they are using.

Imagine if, as part of your ETL procedures, it were possible to predict new values for missing data points. That would help reduce the amount of revenue allocated to the ever-popular “unknown value” bucket. Imagine if it were possible to actually validate existing values to determine if they were in fact correct or, as happens most of the time, just garbage values. This elevates the whole ETL process to a whole new level and adds real business value.

By using predictive analytics, it is possible to automate the discovery of patterns using functions that can sweep through

tables and identify previously hidden patterns. An example of pattern discovery is the analysis of retail sales data to identify seemingly unrelated products that are often purchased together. This information can be used to drive the creation of data-driven hierarchies that allow business users to directly monitor key buying patterns over time.

Other pattern discovery problems include detecting clusters within a column that can then be used to drive completely new hierarchies and drill paths.

Succeeding with ETPL

The following are key requirements for successfully extending existing ETL processes to include a new predict stage:

1. Integrated database engine
2. Industry standard APIs
3. Extensible ETL

Integrated Database Engine

To leverage the opportunities of ETPL, you must start with an integrated database engine that supports both storage and analysis of relational, spatial, multi-dimensional models, predictive analytics, and optimized ETL capabilities. Cross-fertilization of analysis then comes as standard, and because these objects are stored directly within the database, all the processing associated with these features

(such as predictive analytics) can be performed directly within the ETL process as close to the data as possible.

Industry Standard APIs

The industry standard for interrogating and analyzing data stored in an RDBMS is SQL. By providing a SQL and PL/SQL interface to key features such as ETL and predictive analytics, it is possible to seamlessly integrate these powerful features using industry standard languages.

Extensible ETL

The ETL framework must be extensible and make use of the previous two key features. By using industry standard APIs to generate ETL scripts, additional capabilities such as predictive analytics can be integrated seamlessly. By leveraging an integrated database engine, you can safely implement ETPL today, without stretching your existing batch windows to the breaking point.

Delivering Intelligent Data Warehousing

The key to adding value within a data warehouse is to do as much as possible, as close to the data as possible, as early in the process as possible. This is only possible if your database vendor is ETPL enabled. Only ETPL can help you deliver *i*DW. ■

The key to adding value within a data warehouse is to do as much as possible, as close to the data as possible, as early in the process as possible.

MicroStrategy

What are the most effective approaches to self-service BI?

Every day, business people are challenged to uncover information to make timely, informed decisions. While the IT department can provide assistance, the most expedient option is self-service BI.

BI self-service is evolving rapidly. One model offers “what you see is what you get” (WYSIWYG) design interfaces over the Web. With WYSIWYG, business users design and refine reports using familiar skills, similar to PowerPoint and Excel.

The latest model of BI self-service does not require users to design reports. Instead, users surf the data warehouse using “drill-anywhere” capabilities. As users intuitively click on rows, columns, or cells of a report, ROLAP-based BI systems create new SQL and reports. The data can be shared or saved for future use.

Analyst Viewpoint

Business-side engagement and training are key first steps that can help self-service BI live up to its potential. Detailed input from business requirements analysts is essential to ensuring that a self-service BI solution meets the needs of business analysts. Training should not be a mere checklist item, but a prime opportunity for IT to sell business users on the value of self-service BI and gather their feedback. As experience has shown, force-feeding an unfamiliar BI tool can result in user dissatisfaction and subpar adoption. In the simplest terms, business users need to find self-service BI rewarding—even fun—for it to succeed. Business input, collaborative training, and user-friendly WYSIWYG tools are all ingredients for successful self service.

Request more information about MicroStrategy

Netezza Corporation

Power and cooling costs are escalating. What data warehouse options can improve data center efficiency?

Netezza data warehouse appliances were developed with data center efficiency in mind. In one appliance rack, over 100 intelligent storage nodes provide MPP performance by attaching a CPU to each drive. Yet each node consumes less than 30 watts, so the power and cooling requirements for a full-rack system are about 4,000 watts and 12,000 BTU/hour. Compare this with alternative systems, including blade servers, which can consume as much as 600 watts per blade. A rack of 32 blades would require over 19,000 watts and 64,000 BTU/hour of cooling. The efficient design of an architecturally integrated data warehouse appliance offers an excellent solution to address escalating power and cooling costs.

Analyst Viewpoint

Energy efficiency is fast becoming a key consideration for data warehousing and IT as a whole. In fact, by a 417-4 vote in July 2006, the U.S. House of Representatives approved legislation urging IT managers to “give high priority to energy efficiency” when purchasing servers. The good news is that suppliers of data warehouse appliances and conventional servers are making significant strides in engineering energy-efficient systems. Data warehouse administrators will find a number of intriguing options in the marketplace. And with energy prices increasing and fuel supplies dwindling, a transition toward “eco-responsible computing” can reduce energy costs by thousands or even millions of dollars a year for large organizations.

Request more information about Netezza Corporation

Oracle

How is the role of BI evolving in an enterprise?

Enterprises today are looking at BI as a solution to enable intelligent interactions and business process optimization across their entire organization. As a result, the role of BI is evolving to enable pervasive use across the enterprise, empowering users with relevant, actionable, up-to-the-moment business insight that is embedded into users' daily activities. Look for BI vendors who can be your strategic partners and help you gain fastest time to value with a comprehensive, best-of-breed solution encompassing data warehousing, BI platforms, and BI applications.

Analyst Viewpoint

Several overarching trends mark the evolution of enterprise BI. **1)** BI is gradually transitioning from standalone ad hoc systems toward an enterprisewide discipline, including "operational BI," that provides visibility across multiple business units and geographic locations. **2)** Real-time BI is being enabled with on-demand and federated data integration systems that can feed updated data to BI systems. **3)** Organizations are paying greater attention to data at a foundational level, with master data management and data quality initiatives that support BI's ideal of a single version of the truth. **4)** BI is slowly expanding beyond conventional data sources to enable analysis of unstructured data in Web pages, e-mails, word processing documents, and more.

Request more information about Oracle

Sybase, Inc.

Why purchase and manage separate servers, storage, databases, and tools to support a data warehouse when I could purchase an appliance for my business intelligence initiative?

Most organizations are struggling with the total cost of ownership (TCO) of their existing BI platforms. And, while the appliance approach to data warehousing sounds good and will meet the needs of a select group of organizations, most companies find the appliance is not flexible enough, does not leverage their existing systems and hardware, and will ultimately cost more to own over the program's lifetime. For many organizations, a better choice is to integrate their existing applications while leveraging their current hardware to build out a complete BI stack for data warehousing.

Analyst Viewpoint

There's no question that the appliance phenomenon has prompted a reevaluation of the data warehousing price/performance equation, particularly as DW organizations continue to grow data volumes and expand user populations. Appliances, now available at less than \$20,000 per usable terabyte, are an intriguing option for many organizations. None of this, however, spells doom for the conventional best-of-breed approach to DW. Most adopters have deployed appliances to support "large data marts" as they size up price/performance, flexibility, and long-term ROI. Ultimately, of course, data warehousing success doesn't depend on hardware and software alone—but on the skill and productivity of DW developers and administrators. One can, after all, crash a tricycle.

Request more information about Sybase, Inc.

Syncsort Incorporated

My company sells products over the Web, generating high volumes of transaction data. One record per transaction lists the time of sale, dollar amount of the sale, and so on. How would I determine statistics such as the hourly minimum, maximum, and average sales?

Statistical analysis of transaction data can be simplified by using a data management solution that provides minimum, maximum, and average operators, and allows grouping on individual components of the date/time fields. In this example, the aggregate task option would be chosen and the transactions would be grouped by the hour component of the time-of-sale field. Then the minimum, maximum, and average of the sales amount field would be retained. Advanced aggregation functions allow for great flexibility and speed in crunching business analytics, yielding more precise information more quickly.

Analyst Viewpoint

A decade ago, it might have taken a couple of analysts two days of spreadsheet number-crunching to generate such multi-source aggregations. Today's data manipulation tools support a broad range of data types and offer fairly advanced functionality for filtering, joins, summarization, and generation of user-defined statistics. High-performance parallel processing and deployability in distributed network mainframe, UNIX, and Windows environments have helped these tools earn a place in the enterprise data management hierarchy. The scope and complexity of an organization's need to generate certain statistics usually determines whether an Excel client is adequate, or a more costly enterprise solution is warranted.

Request more information about Syncsort Incorporated

XLCCubed, Ltd.

How can I avoid creating "spreadmarts" with Excel?

Spreadmarts are typically a by-product of a failed BI solution, not a failure of Excel. Analysts will create and use spreadmarts if their current solution lacks data or power. Two root causes of spreadmarts are 1) poor business models that lack data and 2) inferior analysis tools. Both will result in analysts inevitably exporting data to Excel—preferring its power—and being forced to acquire data from other sources and store it themselves. When the business models are accurate and Excel is connected directly and dynamically to the source data, spreadmarts will not be created.

Analyst Viewpoint

TDWI research has found that organizations have on average 28.5 spreadmarts that they want to consolidate. While consolidation would help transition toward a "single version of the truth," it's often impractical—and counterproductive—to force analysts to abandon Excel for a standard BI tool. For many organizations, the ideal solution is to enrich and integrate Excel in an enterprise information architecture. Sophisticated Excel add-ons enable organizations to hard-wire Excel into server-based data sources to 1) minimize spreadmart proliferation, 2) improve enterprise data consistency, and 3) allow users to continue using their preferred tool while providing them with the richer ad hoc analysis and reporting available in third-party Excel add-ons.

Request more information about XLCCubed, Ltd.

The Real Value of BI Search

By Wayne W. Eckerson,
Director of Research and Services, TDWI



I must admit I'm a little skeptical about BI search. Ever since Google launched the Google Search Appliance with the OneBox for Enterprise interface to enterprise applications, BI vendors and others have been falling over themselves to partner with Google. The reasons are obvious. Who wouldn't want to team up with the hottest brand in high tech? And who wouldn't want to make information available to end users via the simplest and most ubiquitous interface on the Web?

BI search via Google OneBox makes it very easy for casual business users to search reports, files, records—in short, any object within an enterprise application that interfaces with Google OneBox. (According to Google's Web site, Information Builders, Cognos, SAS Institute, Oracle, Employease, Salesforce.com, and NetSuite have formed OneBox partnerships, with others on the way.) With the OneBox interface, Google's partners can customize the search results to deliver dynamically generated reports and apply security so individuals see only the results they are authorized to view.

Symptom of a Problem. Despite its obvious advantages, BI search is a symptom of a problem. BI search is great when you don't know what you're looking for, or can't find what already exists. But if the

average business user needs a search tool to find basic information or reports about the processes they manage on a daily basis, there is something wrong with the business!

Smart organizations already know exactly what information business users need to manage core business processes. And they set up their information systems to deliver these insights to users on a right-time basis using dashboards and scorecards. These organizations do not want users to waste time searching high and low for information that should already be at their fingertips.

Case in Point. Several years ago when OLAP and ad hoc query tools were the rage, I was surprised to find that the head of marketing at a major casino with a reputation as a leading adopter of data warehousing and BI tools insisted on delivering static reports to his field marketing managers. His rationale was, "We know what information our managers need to run their operations, so we give it to them. We don't want them wasting time submitting queries."

Since then, I've believed that it's better for organizations to push information to users rather than force them to find it. In well-run organizations, business users come to work in the morning, log in to

their computers, and see a dashboard and/or scorecard that displays key metrics that show how their performance stacks up against strategic objectives and goals. After one glance, they know what needs immediate attention and can drill down to shed light on the issue and take steps to address it.

The Pendulum Swing

Companies that rely on BI search as a next-generation ad hoc query tool are missing the point. The pendulum is swinging away from BI self-service, and moving instead to the customized delivery of information. While BI search is a great ad hoc query tool for general business users, do you really want them submitting ad hoc queries all day? Wouldn't it be better to give them role-based reports, dashboards, or scorecards that deliver all the information they need to do their jobs at a glance or within several clicks?

The Role of BI Search. Nevertheless, BI search will fill an important gap in every organization's BI portfolio. For one thing, not all organizations understand their business processes well enough to deliver information and insights proactively via dashboards and scorecards. In addition, some organizations compete in highly volatile markets where strategies, plans,

Perhaps the best use of BI search is not for BI, but as a virtual data integration infrastructure that ties together structured, unstructured, and real-time data behind an intuitive, keyword interface.

processes, and roles change rapidly, placing a premium on information discovery rather than information delivery. Also, business users play many roles within their organizations. In some capacities, they must manage well-defined processes that have fairly static information requirements, but in others, they may engage in more entrepreneurial activities that require ad hoc query or search capabilities.

Finally, business often moves so fast that we lose things quicker than we can find them. A BI search tool can accelerate “time to recover” what we’ve lost or misplaced by moving too fast, doing too much, and trying too hard (although I can almost hear Peter Drucker saying that these are symptoms of an organization out of control!).

Poor Man’s Data Integration

Perhaps the most overlooked benefit of BI search is its ability to substitute as a data integration service. Rather than purchase complex middleware and hire expensive systems integrators, companies can implement Google OneBox to pull together records and objects from multiple systems for little money, including both structured (databases) and unstructured (files, text, images, video, etc.) sources. (Google Search Appliance starts at \$30,000 to

index 50,000 documents.) Although Google OneBox can’t move records between applications, and users have to manually “piece together” information from potentially dozens or hundreds of search results, it may be sufficient to meet certain data integration requirements. It is a cheap way to bridge the formidable gulf that currently exists between structured and unstructured data.

Perhaps even more interesting, data integration vendors such as iWay Software (a subsidiary of Information Builders) can interface their application and data adapters to Google OneBox, feeding Google with transactions as they occur. By indexing events on the fly, Google OneBox can deliver decision makers up-to-date information in response to keyword searches. So, when tied to key middleware components, Google OneBox can support real-time decision making.

Conclusion

There’s been a lot of talk about the Google-ization of BI. Some of this is hype, as BI vendors clamor to jump on Google’s bandwagon. But some is legitimate, since the Google OneBox for Enterprise finally provides an ad hoc query interface that general business users find easy and intuitive. While BI search should never be used as a primary BI tool, it makes a

wonderful supplement to the information organizations should be delivering to users on a timely basis via dashboards and scorecards. Perhaps the best use of BI search is not for BI, but as a virtual data integration infrastructure that ties together structured, unstructured, and real-time data behind an intuitive, keyword interface. ■

Wayne W. Eckerson is the director of research and services for TDWI. He has 17 years of industry experience and has covered data warehousing and business intelligence since 1995. Eckerson is the author of many in-depth reports, a columnist for several business and technology magazines, and a noted speaker and consultant. His book, Performance Dashboards: Measuring, Monitoring, and Managing Your Business, was published by John Wiley & Sons in 2005. He can be reached at weckerson@tdwi.org.

Solving Old Problems with EIM

By Frank Dravis, Vice President of EIM Strategies, Business Objects

Have you ever heard your IT staff say, “I just need access to that data,” or, “Providing you that data would be easy, if I knew how to find it”? Helping business analysts to access and use data stored in proprietary applications can be a challenge. The application confines the capture data to that specific operation and restricts it to that environment. Take SAP, PeopleSoft, or Siebel, for example. These applications work well for their intended purposes, but what happens when you want to access and process the data in ways that aren’t supported by the vendor?

Data profiling is a case in point. Suppose a marketing department is reporting substantial duplicate contact records in its CRM system. Many of the duplicates have the same last name and phone number, but there are differences in the first names, like Bob versus Robert, or differences in the firm names, like APCC versus American Power. These variations produce duplicate contacts, which creates the need to find and analyze the domain errors in each field—data profiling and cleansing tasks.

Connecting a profiling solution to the back end of a CRM system, or any proprietary application, to analyze the underlying data can be problematic because each application has its own proprietary “business layer language.” Coding against this layer to perform near-infinite analysis tasks is a significant undertaking with one application. Supporting multiple applications is even more daunting. A profiling tool can connect directly to the back-end database of these applications because they usually use DB2, Oracle, or Microsoft SQL Server. However, the back-end physical data model usually has no easily discernible relationship to the logical user view displayed by the GUI.

This is where the synergies of functionality across an EIM (enterprise information management) framework come into play. Organizations will deploy

ETL, metadata management, data federation, data quality, and other solutions found in an EIM framework to build out their IT infrastructures and allow them to connect disparate data sources to their BI, data warehouse, or CRM operations. A common practice is to build data marts from extracts of proprietary applications. Some packaged solutions allow you to quickly extract domain-specific data sets from the application and load them into a predefined data mart that matches the domain context; for example, a data mart for product data from SAP. The benefit is that these ready extracts, or “rapid marts,” require only limited knowledge of the proprietary application.

With these rapid marts, the data profiling or cleansing solution can access the data easily because the tables and columns have been exposed in a user-friendly fashion. The profiling software can even provide templates that automatically connect to the rapid mart, run a standard set of analysis tasks, and report on the findings with little user setup. Data stewards can then profile the data in the pro-



proprietary application without specialized knowledge of the back-end data model. Defect reports are easy to generate; the field mappings are provided, so IT knows what specific data fields in the application need to be cleansed. In other words, the drudgery of having to wade around in 7,000 tables of the back-end database disappears. Moreover, if the data in the rapid mart is needed for downstream operations, it can be cleansed at that point without affecting the application.

By combining and linking the functions within an EIM framework, new and innovative approaches to old problems become possible. These synergies surface only when data managers look across the breadth of EIM and consider unique ways of supporting one capability with another. ■

By combining and linking the functions within an EIM framework, new and innovative approaches to old problems become possible. These synergies surface only when data managers look across the breadth of EIM.

As Business Intelligence Becomes a Strategic Imperative, Is Your Infrastructure Ready to Handle It?

By Ellen Rubin, Vice President of Marketing, Netezza Corporation

Savvy, market-leading companies across a range of industries are “competing on analytics” to dominate their markets. These companies understand that business intelligence has evolved from tactical reporting to a strategic imperative and they must fully leverage it to stay ahead of the pack. Ironically, as BI is growing in importance across the enterprise, it’s also becoming harder to achieve success. What are the hurdles to successful BI initiatives? How does your infrastructure need to evolve to keep pace?

Hurdles to Successful BI

As companies broaden their BI initiatives, they must upgrade their capabilities to handle increasing enterprise pressures, including:

- **Exploding data volumes and shrinking data latency tolerances.** As businesses collect huge amounts of data from new sources and applications, they’re also under increasing pressure to know now, not tomorrow, what intelligence is locked in the data.
- **Increasing analytic sophistication.** Businesses are no longer interested just in what has happened, but also in what will happen, so they need to run ad hoc analyses on detailed data, review the results, and continue to iterate.
- **Expanding user community.** As BI becomes increasingly strategic and pervasive, more people (internally and externally, from power users to more casual ones) need to access the data.

But, as noted in a *DM Review* article on OLAP,¹ one of the biggest hurdles to effective BI is slow query performance caused by the limitations of the underlying analytic infrastructure. Traditional data warehouse systems can’t keep pace with

One of the biggest hurdles to effective BI is slow query performance caused by the limitations of the underlying analytic infrastructure.

today’s BI needs—they weren’t designed to quickly deliver complex analytics on terabytes of data. They’re expensive to scale, complex to administer, and time-consuming to deploy. These limitations are diametrically opposed to the growing demands of BI.

Evolving Your Infrastructure

The Netezza Performance Server data warehouse appliance offers a revolutionary approach to data warehousing. It was designed specifically for the challenges of today’s BI:

- **Purpose-built for performance.** Coming from a single vendor, the architecturally integrated database, server, and storage are designed specifically for high-performance data warehousing. This includes dedicated hardware for processing large data volumes faster than any other data warehouse solution in the market.
- **Flexible and scalable.** The appliance grows modularly with business needs.
- **Simple to use.** Easy to install, deploy, and maintain. A true data warehouse appliance requires no tuning, indexing, partitioning, or aggregations.
- **Low cost.** Significantly less costly to own and maintain—even for large EDW implementations of 100 terabytes or more—leading to a lower overall total cost of ownership.
- **Enterprise compatible.** Highly available, with “plug ‘n play” integration and installation in just hours. Appliances have standards-

based interfaces and are fully integrated with all major BI and data integration vendors.

By leveraging the Netezza data warehouse appliance, companies can run existing analyses faster and more deeply, but even more important, they can run new, previously impossible analyses to drive business growth.

Proving It Out

Netezza’s customers are major companies across data-intensive industries that are now leveraging BI for competitive advantage—because they can. One of Netezza’s large financial services customers is now saving \$100 million annually on postage because they can run much more targeted customer segmentations for direct mail promotional offers. A large online business can now quickly understand the effects of changes made to their site, and make immediate adjustments to avoid lost revenues. And with Netezza, a leading retailer is realizing a several-fold reduction in data load times and can now move from weekly to daily updates to its sales transactions data warehouse, allowing the company to reduce cycle times in identifying shifting sales trends and deliver significant bottom-line impact. Netezza’s customers understand that using the innovative Netezza data warehouse appliance as part of their infrastructure is key to the success of their BI initiatives and integral to competing on analytics. ■

¹ Nigel Pendse, “What’s New and What’s Not?” *DM Review*, February 2005.

High-Performance Technologies Add Up to Successful Business Intelligence

By Susan Garretson Friedman, Senior Technical Analyst, Syncsort Incorporated

Growth rates and new opportunities for all kinds of businesses, in particular banks and other financial institutions, mean they have to process skyrocketing amounts of data. In financial institutions, the data glut is further complicated by the need to comply with legislation such as the Basel II Accords, Check 21, Sarbanes-Oxley, and the USA PATRIOT Act or risk running afoul of government regulations. More financial institutions are also realizing that they need fast and accurate intelligence from their data to stay competitive in today's real-time business world. Data management, and data governance, have become priorities. In 2005, according to a Knightsbridge Solutions white paper, these pressures led to financial institutions making major investments, totaling more than \$1 billion industry-wide, in basic data infrastructure and data governance. Much of that money was spent trying to improve access to, and the quality of, real-time business intelligence.

The consulting group Gartner has been quoted in *DM Review* as recommending that organizations focus on the following financial reporting criteria:

- Collecting financial data
- Making financial details more accessible
- Drilling down on accounting reports
- Highlighting key analysis areas based on tolerance and financial metrics
- Segmenting reporting into significant elements
- Enabling workflow
- Offering frequent flash reporting
- Instilling a financial management mindset within the organization

Business intelligence (BI) offers enterprises the ability to meet all those criteria as well as the requirements for competitive analysis.

For processing data for BI analytics, the usual go-to software includes various ETL and data warehousing solutions. However, we have discovered that these tools do not always yield the peak performance that institutions need to maximize their intelligence gathering and minimize the processing time it takes to get it. Often an additional tool can contribute to speeding up the processing times of high volumes of data. In particular, the higher-level functions, such as high-performance joins, aggregations, string, time/date, arithmetic, and user-defined functions, can help enormously when combined with existing database or ETL software, or in speeding existing company-generated processes.

A Quick Overview

High-Performance Aggregation

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.

High-performance aggregation simplifies the creation, administration, and execution of aggregation jobs. It summarizes data much faster than other aggregation methods such as C programs, SQL statements, or third-party, multi-purpose data warehouse packages. It provides the necessary flexibility to select the best aggregates for optimizing query performance.

High-Performance Joins

High-performance joins significantly improve the efficiency of preprocessing, retrieval, and updating in a dimensional data warehouse, providing the following significant benefits:

- Optimize data preparation
- Improve query performance
- Reduce the quantity of data that must be processed in response to a query
- Speed lookups and application matching
- Retrieve and summarize information more efficiently
- Minimize storage and throughput requirements
- Reduce elapsed time of change data captures and delta processing

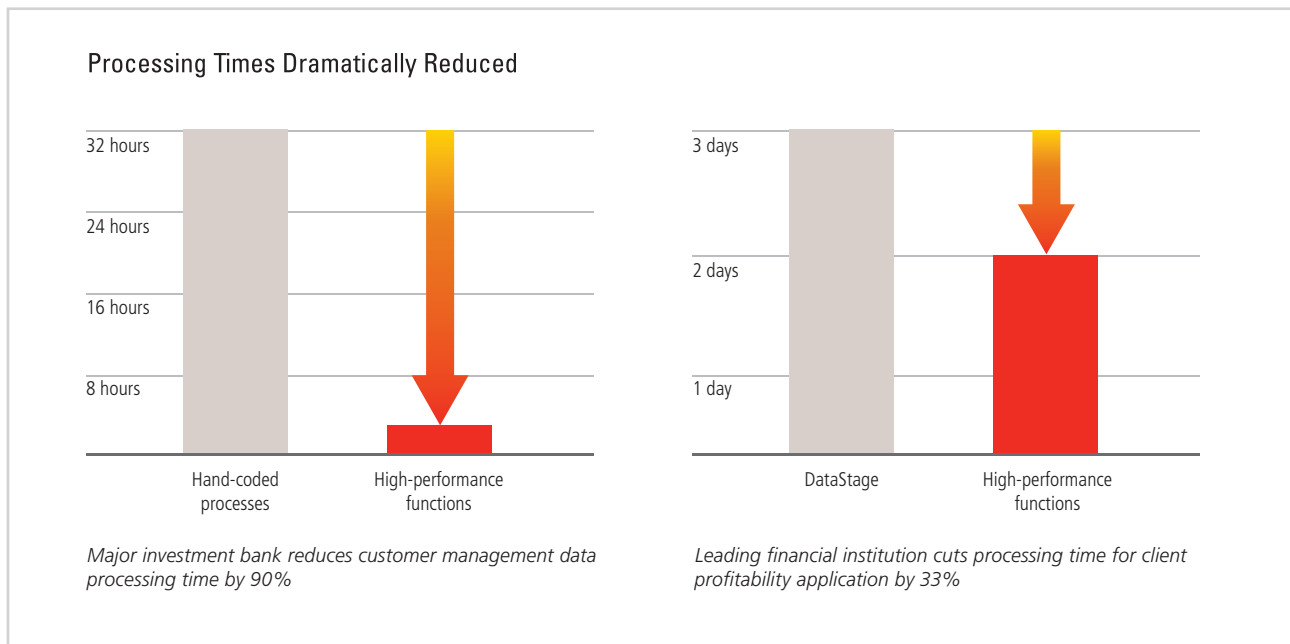
Additional Functions

Additional data transformation capabilities include arithmetic functions, string functions, date/time functions, and user-defined values. User-defined values allow easy definition of tasks that enable powerful data transformations and perform complex processing.

Real-Life Examples

A number of our customers have found that adding these advanced, high-performance functions yields excellent results. One major investment bank experienced a 90 percent reduction in their customer management data processing time.

They wanted to improve the performance of a lengthy product/customer data management process that originally involved five major procedures, all written in Java. Legacy Java-based applications were consuming vast amounts of system



resources and time due to recent increases in data volumes. The bank's applications needed to be upgraded in order to reduce elapsed time for routine customer management processes. The processes included data extractions, joins, merges, reorganizations, and loads in a networked UNIX environment. The bank's existing solution, using programs written in Java, took 30–32 hours to complete.

The solution: The Java procedures were converted to high-performance features in stages, eliminating the need for a separate database load step and allowing all files to be tabbed. Increased time savings were achieved as the additional stages were implemented: Total reduction in elapsed time was over 90 percent, from 30–32 hours to just 2–3 hours. The bank can now perform its routine customer management processes faster and more frequently.

At another leading financial institution, a specific group is responsible for processing customer data as part of a customer profitability project. The group designed a client profitability application for the project that examines customer account and product data in order to determine how much (and at what rates) the company's products have earned for

their investment clients. The application uses a Sun E10K system with 24 CPUs and 20 gigabytes of memory running Solaris 2.6.

As part of the application, the company keeps the current year's data along with historical data from the two previous years in the operational data store (ODS). The ODS goes through about 50 billion calculations a month using Java, and then the information is stored in an Oracle 8i database. The data warehouse loads the output tables of that ODS, which contains all of the financial data. They then extract it through PL SQL into time periods that provide the financial data by customer account and product in edited numeric (EN) format.

The data from the time periods is processed using DataStage, which then outputs financial files by time periods as well as by product, customer, organization, and other types of keys. There are 36 files for regular time periods that are approximately 1 gigabyte each. There are also five summary time period files that are approximately 1.8 gigabytes each. The company then needs to create aggregates and order the data. The group responsible for the application found that DataStage's aggregation and ordering features did not

provide the performance and functionality that they needed. With DataStage alone, the process was taking 72 hours.

To improve this situation, the company selected high-performance functions to carry out the aggregates and ordering. Once the aggregates are complete, the database is brought down. The database administrator runs several statistics with the aggregated files and other DataStage files. The data is then loaded into the data warehouse and different divisions from around the world query the information using Oracle Reports and Oracle Discoverer. When they run the ETL process with the high-performance technology, the old data is still up, so they're able to perform queries against the processor while running their ETL process. They were able to reduce the elapsed time to just 48 hours—an improvement of 33 percent.

Conclusion

These are just two examples of the wide range of benefits that companies, and banking enterprises in particular, can reap from using high-performance joins, aggregates, date/time, arithmetic, and user-defined functions—and they show that using high-level functions really adds up to success. ■

Enterprise Information Integration: A Technology For Providing Integrated Views

By Ashok Swaminathan, Director, Product Management—Sybase Data Integration, Sybase, Inc.

For organizations to extract value from multiple data sources quickly and efficiently, they must be able to integrate current data from heterogeneous sources. Three methods for developing this capability include:

1. Accessing the data sources directly.

However, keep the following in mind:

- Application developers must know what data to access and where and how to access it. Writing access logic (versus business logic) takes time. And making changes in data sources (adding or modifying tables or columns) means that the consuming applications must also be modified.
- Applications are tightly coupled with application servers and databases. Data access logic is therefore typically buried in the application logic, leaving little scope for reusability.
- Data owners have no easy way to see what parts of their data are used by applications, and no convenient way to control those applications' access to that data.

2. Using ETL technology to populate a data warehouse. Extract data from the operation store (usually through batch processing), transform, and make the data suitable for consumption by the decision support processes. The target is a data warehouse or operational data store. This approach has advantages, but is not optimal for current access to heterogeneous sources of data.

3. Using enterprise application integration (EAI) technology to integrate the business processes of two or more applications via their APIs. EAI provides process-to-process integration and helps execute transactions and process flows. However, it's not meant to process

complex queries across data sources, aggregate the results, and serve them up as a result set.

While these technologies each serve different purposes, EII is best when users want to create unified views of data without making copies of the source information.

Enterprise Information Integration (EII)

EII solutions provide a means for integrating data in real time, allowing standardized access to the integrated data through a single, federated data layer. EII technology is ideal for applications that require current data that is selectively integrated from many sources:

- There is no need to rewrite applications to integrate a new data source or if the physical location of the data sources changes. New sources can be incrementally added to the federated data layer without affecting existing applications.
- Developers can find queries or services that have already been developed, and if appropriate, reuse them rather than recreate them. EII provides significant savings because developers can focus on the business logic rather than the access code.
- Data owners can see what aspects of their data they've made available to consuming applications. They can control that availability in terms of access authorization and the load on their data sources. Plus, they can track how this data is used.

Typical Uses of EII Technology

EII's distinguishing characteristic is its ability to access and integrate current data from heterogeneous sources. How? A virtualized data federation layer delivers data to applications in real time from original sources. Various vertical segments increasingly use EII for business decisions:

1. **Banking.** Banks provide multiple lines of service: checking, home equity loans, brokerage services, etc. Banks want a complete view of customer activities across lines of business from multiple divisions and applications. Most banking data is in silos, making EII a natural choice for integration.
2. **Securities Trading.** As hedge funds trade across asset classes, traders need a complete view that incorporates up-to-the-minute data. Typically, different applications generate trading data, and different databases store it. EII allows integration from different sources, providing complete trading information across asset classes.
3. **Single Customer View.** As organizations roll out services, customer service representatives need complete views across all lines of business. For example, insurance companies are adding services besides life, automobile, and homeowners insurance. Service representatives must access all relevant data. EII technology is ideal for this application.
4. **Federal Government.** Certain laws and regulations make sharing information difficult. However, Homeland Security applications must integrate the data government agencies collect. Aggregating and copying data into a central repository won't work. Agencies need to expose particular data elements, stipulate uses, and make specific data available to other agencies. EII meets this need.

EII-based data integration boasts a powerful new kind of enterprise technology that uses data federation. Visit: www.sybase.com/products/developmentintegration/avakieii or e-mail us at infobi@sybase.com. ■

Appliances—Data Mart or Enterprise Data Warehouse?

By Stuart Frost, CEO, DATAlegro, Inc.

Apliances 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.

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 expensive EDW upgrades. In addition, the specialized nature and advanced technology of

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

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? ■

Master Data Management: Consensus-Driven Data Definitions for Cross-Application Consistency

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

Overview of Master Data Management

For many people in IT, this decade is all about integration. It's about integrating customer data, integrating application silos, integrating data for BI purposes, integrating with partners, integrating with governments, and integrating through new interfaces like Web services. As if the list weren't long enough, the practice known as *master data management* (MDM) has become a candidate for even more integration work.

Hence, a lot of data-management professionals and their business counterparts are asking: "What is master data management? Why should I care? Why is it imperative now?" This report seeks to answer these and related questions.

What Is Master Data Management?

We need to build up to an answer to this question by describing (in procedural order) the technical components and best practices that are subsets of master data management:

- **Master Data.** Master data (sometimes called reference data) consists of facts that define a business entity, facts that may be used to model one or more definitions or views of an entity. Entity definitions based on master data provide

business consistency and data integrity when multiple IT systems across an organization (or beyond) identify the same entity differently.

- **Business Entity Definitions.** Today, the business entities modeled via master data are usually customer, product, or finances. But master data can define any entity, like employee, supplier, location, asset, claim, policy, patient, citizen, chart of accounts, and so on.
- **System of Record.** Common across the many approaches to MDM is the creation (or selection) of a system of record (sometimes called a trusted source). The point is to establish a central, authenticated master copy from which entity definitions (and perhaps physical data, too) are propagated among all IT systems integrated via the MDM solution.
- **MDM Hubs, Master Files, and Application Databases.** The system of record can take many forms. Many users build a central database (like a data warehouse or operational data store) as a hub through which master data, metadata, and physical data are synchronized. And some hubs are simply master files or master tables that collect and collate records. Sometimes a pre-existing application (typically for ERP or CRM) has the needed definitions already, so it's selected as the system of record and the basis for entity modeling elsewhere.
- **Master Data Integration.** Regardless of the technology approach, the goal of the system of record is to provide a central mechanism for collecting and sharing consistent definitions, usually across otherwise unrelated IT systems. Obviously, this requires technologies and best practices for system integration, data integration, and application integration. Hence, many technical users consider MDM to be an integration practice, enabled by integration tools and techniques for ETL, EAI, EII, and replication. When the system of record is a hub that connects many diverse systems, multiple integration technologies may be required, including newer ones like Web services and service oriented architecture (SOA).
- **Master Data Management.** Drawing from the above definitions, we can now say that MDM is the practice of defining and maintaining consistent definitions of business entities, then sharing them via integration techniques across multiple IT systems within an enterprise and sometimes beyond to partnering

Enterprise MDM and data flow

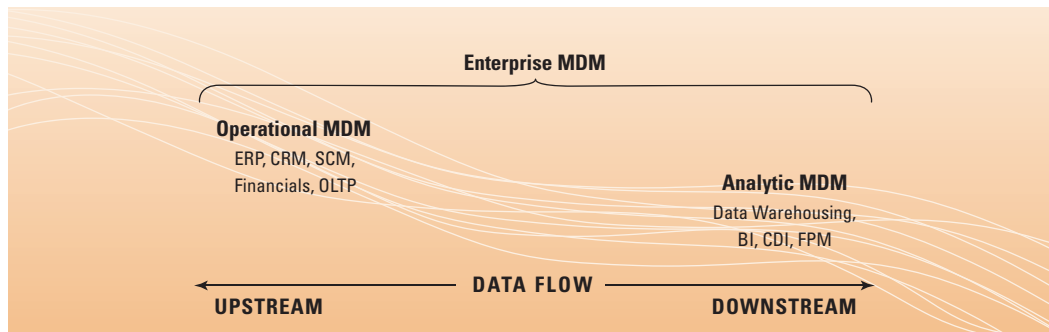


Figure 1. MDM solutions can be upstream or downstream relative to the flow of data.

companies or customers. More simply put: MDM is the practice of acquiring, improving, and sharing master data.

MDM Solution Types. MDM solutions in most organizations are built into or closely associated with a larger application, though some span multiple applications. Due to these associations, MDM solutions fall into three broad categories (see Figure 1):

- **Operational MDM.** Upstream in the general flow of data, one or more MDM solutions are built into and/or used to integrate operational applications for enterprise resource planning (ERP), customer relationship management (CRM), supply chain management (SCM), financials, and so on. Since the applications don't change frequently, master data and entity definitions don't either. Furthermore, these applications and their business processes support transactions and sometimes operate in real time, so operational MDM must, too.
- **Analytic MDM.** Downstream in the data flow, data warehousing has long involved some form of MDM, because of the balance between tracking data lineage (to ensure you have the right data) and repurposing data to

create new structures (like aggregates and time series). Entity definitions change often due to data discovery, analytic business modeling, and other iterative practices. Analytic MDM is also seen in practices that resemble data warehousing, like customer data integration (CDI) and financial performance management (FPM).

- **Enterprise MDM.** Today, MDM is practiced mostly in isolated silos or with a short list of applications that don't step beyond the bounds of either operational MDM or analytic MDM. However, some organizations have moved to the next level with *enterprise MDM*, which is an autonomous infrastructure that can integrate master data across multiple IT systems and businesses. Spanning the whole data flow is daunting, because enterprise MDM must satisfy the diverse requirements of both operational MDM and analytic MDM. Yet, enterprise MDM is a worthy goal, because it extends beyond IT silos and organizational boundaries the general benefits of MDM, namely: well-designed entity definitions applied consistently.

Why Should We Care about Master Data Management?

The main reason that technical and business people should care about MDM is

the long list of problems that occur when it's ignored. Somewhat less compelling is the short list of benefits that come from improving master data and its integration across IT systems.

The Problems of Poor Master Data.

In an Internet-based survey that TDWI ran in mid-2006, a whopping 83% of respondents reported that their organizations have suffered problems due to poor master data. (See Figure 2.) Hence, the vast majority of users attests that MDM problems are real, numerous, and severe.

Poor-quality master data creates problems mostly within data warehousing and BI—but also outside it. (See Figure 3.) The top three problems relate directly to data warehousing, namely inaccurate reporting (81%), arguments over which data is appropriate (78%), and bad decisions based on incorrect definitions (54%). Others are general data management problems that sometimes impact data warehousing, like data governance and stewardship limitations (53%), limited visibility for data lineage (52%), and no understanding of master data homonyms (46%). Other problems cited by survey participants are business problems, like poor customer service (35%), inefficient marketing (32%) or purchasing (18%), and new products delayed (17%).

Has your organization suffered problems due to poor master data?

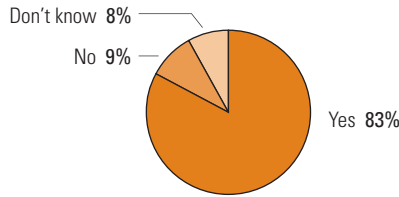


Figure 2. Based on 741 respondents.

What kind of problems?

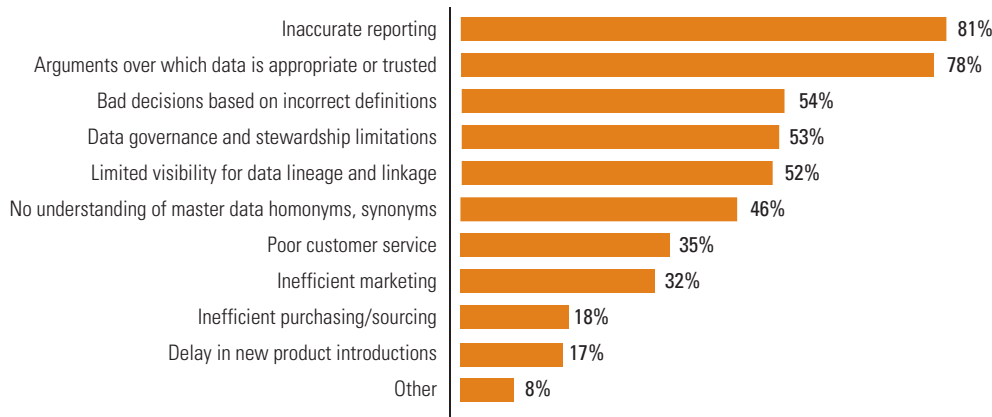


Figure 3. Based on 2,921 responses from 615 respondents.

The Benefits of High-Quality Data.

Though a vast majority of users have suffered problems from poor master data, a much slimmer majority (54%) claims to have derived benefits from good master data. (See Figure 4.) Still, this indicates that benefits are possible and can be identified.

As with master data problems, the benefits likewise relate most strongly to data warehousing and related data management practices, followed by general business issues. (See Figure 5.) Near the top of the list are data warehousing and BI issues, like accurate reporting (75%), better decision making (69%), and easier auditing of information’s origins (47%). Mixed with these are general data management issues, like data quality (76%), consistent definitions (65%), and data governance

(39%). Low-ranked benefits are mostly general business issues, like risk reduction (33%), superior customer service (21%), and supply chain optimization (15%).

Good MDM Yields Good Reports. Users can clearly see the impact of MDM—both negative and positive—on visible products of information management, especially reports of various types. For example, in TDWI’s survey, inaccurate reporting is the leading problem and accurate reporting is the second leading benefit. The survey aside, most of the users that TDWI interviewed pointed to reports as an area where both technical and business people seek improvement via MDM.

Though all reports may benefit from improved MDM, regulatory and financial reports are a hot spot, because they are

scrutinized carefully today and can result in dire consequences when discrepancies are found. In fact, many interviewees admitted that they and their peers live in fear of an audit, and that they feel MDM can help them avert or prepare for such an event. For example, the consistently applied definitions of MDM ensure that reports are populated with correct data, and the data lineage of MDM answers questions in the event of an audit. Similar concerns are seen in the survey, where roughly half of respondents reported “limited visibility for data lineage” as a problem and “easier auditing of information’s origins” as a benefit. Despite its negative impetus, “audit paranoia” is a significant driver for increased efforts in master data management, as well as related areas like data quality, data warehousing, and report design.

Has your organization derived benefits from good master data?

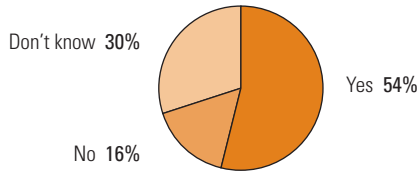


Figure 4. Based on 741 respondents.

What kind of benefits?

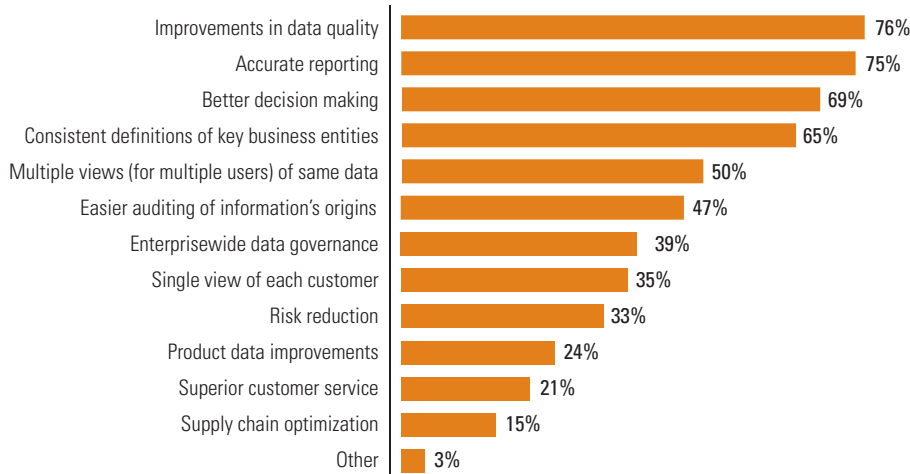


Figure 5. Based on 2,224 responses from 402 respondents.

Why Is Master Data Management Imperative Now?

Trends unique to this first decade of the new millennium have brought MDM to the forefront:

- **Technology Trends.** As an integration practice, MDM is late to the data integration frenzy of this decade, yet still part of it. For many technical users, MDM is an extension of other growing practices, like data warehousing, BI, data quality, and metadata management. In some cases, MDM's central definitions are part of a general trend toward centralized IT.
- **Business Trends.** After Enron and WorldCom, many corporations have entered an age of accountability, where executive actions are driven by audit paranoia, which MDM helps

alleviate. Market consolidation is rampant in some industries (banking, high tech, telco), and MDM can be a useful component in the quick-but-critical integration of customers, financials, and products demanded of a merger, acquisition, or corporate reorganization.

- **MDM Trends.** Although MDM has long existed in a few isolated silo applications, companies now practice it in more silos and with more connections among silos. Forward-looking companies even practice enterprise MDM, with a broad scope across many applications and businesses. As with many data management practices, the general trend in MDM is toward broader and better integrated enterprise scope. ■

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

This article was excerpted from the full October 2006 report. TDWI appreciates the sponsorship of Actuate Software, ASG, Business Objects, DataFlux, Informatica, SAP, Sunopsis, and Teradata.

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

Winners: Best Practices Awards 2006



TDWI's Best Practices Awards recognize organizations for developing and implementing world-class business intelligence and data warehousing solutions. Here are summaries of the winning solutions for 2006. For more information, visit www.tdwi.org/BP06.

CATEGORY

BI/DW ON A LIMITED BUDGET

1-800 CONTACTS

1-800 CONTACTS is the world's largest supplier of contact lenses. No one is more likely to have a customer's exact prescription in stock and ready to ship than 1-800 CONTACTS. The company is dedicated to providing excellent customer service. Orders are placed by phone or via the Web (1800contacts.com). Company headquarters are located in Draper, UT.

In April 2005, the foundation of a new data warehouse was laid and an order activity data mart was completed. This met the basic needs of senior management and key analysts by giving them a consistent view of daily performance, but it did little to leverage the available information to improve daily performance in the call center.

A project was proposed to create a performance dashboard that the agents would be able to see continually, providing them with the information necessary to change behaviors and increase performance. Monthly incentive bonuses would be tied to agents' performance based on a combination of performance measures: closing ratio, average order size, productivity, and quality score. Using relative indices instead of actual values provided an equitable way to compare all agents.

In the deployed solution, a dashboard of gauges and charts shows agents how they are performing in each of the scoring categories. The data is updated every 15 minutes. The dashboard not only shows them how they are performing for the current day, but also compares them to their own performance month-to-date, to the call center averages, to the other members of their team, and to the goals that they set for themselves. The gauges, charts, and indicators give them a quick visualization of their performance without requiring them to page through reports or spreadsheets.

Using indexing of actual scores, where only the agents that worked the same minutes were compared to each other, created a level playing field for comparison. Giving the agents their detailed performance information, in a timely way and in an easy-to-read dashboard, not only established a competitive environment, but also empowered them with the information they needed to improve.

CATEGORY

BUSINESS PERFORMANCE MANAGEMENT

BNSF Railway, Inc.

A subsidiary of Burlington Northern Santa Fe Corporation (NYSE:BNI), BNSF Railway Company operates one of the largest railroad networks in North America, with 32,000 route miles covering 28 states and two Canadian provinces. The railway is among the world's top transporters of intermodal traffic, moves more grain than any other American railroad, transports the components of many of the products we depend on daily, and hauls enough low-sulphur coal to generate about 10 percent of the electricity produced in the U.S. BNSF is an industry leader, Web-enabling a variety of customer transactions at www.bnsf.com.

As part of BNSF's ongoing commitment to operational excellence, the company successfully launched its corporate dashboard in 2005 to provide real-time views of critical operating measures for BNSF's management team. The dashboard accelerates the research process, allowing users to understand information more quickly and enhancing the departments' ability to make informed decisions.

Key to the dashboard's effectiveness is its automated reporting on 40+ significant operational measures. These measures are color-coded to allow viewers a quick assessment of trends and current status. From crew starts to locomotive velocity, from safety measures to revenue and on-time performance, the dashboard has the streamlined ability to drill down into critical operating measures and provide comparisons against historical views. This process is essential to managers' understanding of past performance and their ability to take subsequent action. By displaying clear, factual information, BNSF's corporate dashboard tool assists management in the day-to-day operations of the railroad.

The dashboard demonstrates creative methods of visualization, a "map of the market" view. Users can view the daily performance of BNSF's railroad activities, including safety, train statistics, fuel, car cycle, revenue units, crew starts, and car inventory, among others. On the same screen, there is a geographic display map in which users can select a specific location to obtain detailed information regarding activities at a given locale on BNSF's railroad network.

CATEGORY

DATA GOVERNANCE

IBM

In 2005, International Business Machines Corporation (IBM) had more than 329,000 employees worldwide and revenues of \$91.1 billion. While IBM produces and delivers computer hardware and software services, IBM is also an innovation company.

IBM's Enterprise Business Information (EBI) initiative began in 2002 with a focus on customer information, sales reporting, and governance responsibilities for customer data. Over the course of the next four years, the organization gained data responsibilities in many data areas (ordering, offerings, sales knowledge, business partners, software products, analytics, and enterprise data warehouses) while expanding data governance, first for IBM's supply chain organization and then for the complete enterprise.

With all of these information management responsibilities in a company as large and complex as IBM came the need for more encompassing data governance and stewardship leadership. IBM's solution was an EBI center of excellence (COE).

The EBI COE provides the foundation for the CIO's strategic imperatives by recognizing information as an asset and enabling internal transformation through processes, methods, roles, and services, thus building an integrated business information environment. The EBI COE components are intentionally aligned with the senior executive processes and management teams controlling the direction of all enterprise initiatives and projects. This has elevated the "information health" of enterprise initiatives and projects, ensuring the quality of IBM's enterprise data. The company's ability to integrate data into information and knowledge for business use bears directly into the funding decisions as IBM's transformation activities advance.

From a data governance perspective, the initial components of the EBI COE system (information management framework components, IM management system, and IM stewardship programs) have institutionalized themselves within IBM. The stewardship programs have raised information quality focus across all executives in the business information value chains. The momentum established with these early EBI COE components is carrying forward with an ever-increasing number of EBI COE service requests.

While the EBI COE is still considered a relatively new effort, there have already been tangible savings realized. In the ETL aspects alone, a 20 percent cost avoidance for every \$1 million project is being realized through the use of EBI COE services and use of EBI COE common corporate software.

CATEGORY

ENTERPRISE DATA WAREHOUSING

Royal Bank of Canada

SOLUTION SPONSOR: TERADATA, A DIVISION OF NCR

Royal Bank of Canada is Canada's largest bank as measured by market capitalization and assets, and is one of North America's leading diversified financial services companies. RBC provides personal and commercial banking services, wealth management, investment banking, insurance, and transaction processing services to more than 14 million clients in 30 countries.

As it grew and diversified, RBC accumulated vast amounts of customer and transactional data in divisional product and operating systems. By the mid-1990s, the need for data became a paramount requirement to the firm's marketing, customer service, and risk management areas. Business units began building single-purpose data marts for reporting and analysis.

Data architects within RBC's enterprise information management (EIM) group realized that an enterprise data management solution was essential to provide a unified view of the business and its customers, as well as to control costs, eliminate redundancy and duplication, ensure consistent representation, and improve business agility.

With support from executive management, the EIM group undertook a long-range, incremental program of infrastructure development, beginning with the creation of a staging environment to support data transformation and quality control. This allowed them to impose data management standards and quality control processes at the earliest possible moment, while extracting maximum value and performance from existing data mart investments.

Eventually a small-scale warehouse was implemented to support marketing operations, demonstrate the utility of a common repository, and build support for a true enterprise facility. In 2000 that support was realized, and a major data warehouse/data mart consolidation effort was launched based on the NCR/Teradata platform. Since that time, growth and development have been continuous.

The enterprise data warehouse architecture was designed from the outset to enable scalability, reliability, flexibility, and adaptability over the long term. In particular, the EIM team took a disciplined approach to data structure definition, providing an integrated data and semantic layer that enables efficient management and fast, cost-effective reuse of data. The EDW currently executes more than a million queries and up to 25,000 ETL batch processes every month. It houses more than 30 terabytes of business data, and supports more than 33,000 users.

CATEGORY

GOVERNMENT AND NON-PROFIT

Broward County Public Schools

SOLUTION SPONSOR: IBM

Broward County Public Schools, in Florida, is the sixth-largest school district in the U.S., with a unique urban/suburban mix of students. It is the largest fully accredited district. The county's 108 offices and 281 schools are home to more than 270,200 pre-kindergarten through twelfth-grade students, 75,000 adult education students, and more than 41,900 employees, of which more than 17,000 are teachers. With a budget for the 2004/05 school year of \$4.14 billion, the Broward County School Board must be run as any large company would.

The Broward School District built a data warehouse, deployed a BI solution, and staffed a supporting organization that resembles a market-driven company in the private sector. The data warehouse reporting system has dramatically changed the way Broward educates students and propels them to achieve, far beyond what was previously thought to be attainable. Data and reports permeate nearly every aspect of state reporting, monitoring, class management, test score improvement, and student counseling.

Broward County's robust data warehouse and reporting solution have become the backbone of all data and reporting activities at Broward. Thanks to the Web-based reporting, all users now have quick and easy access to data. With many different types of computers, the Web-based reporting environment is ideal for schools and for users who have varying levels of technical proficiency.

The symptoms of the lack of data and reports were painful to district employees and very expensive for the district. Now, warehouse data is used throughout the district for a wide variety of reporting and student assistance uses, including:

- Grades
- Attendance monitoring
- Exceptional student tracking
- Standardized test score monitoring
- Student health records
- Teacher/class composition
- Staff certification
- Student discipline tracking
- Curriculum planning based on test score performance
- State reports on high school student absenteeism
- State-required reporting on a number of different criteria

CATEGORY

MASTER DATA MANAGEMENT

Nationwide Insurance

SOLUTION CO-SPONSORS: INFORMATICA CORPORATION
AND TERADATA, A DIVISION OF NCR

Nationwide is one of the largest insurance and financial services companies in the world, with more than \$157 billion in statutory assets. Core businesses include domestic property and casualty insurance, life insurance and retirement savings, asset management, and strategic investments. Nationwide provides a full range of products and financial services from auto, fire, life, health, and commercial insurance to administrative services, annuities, mutual funds, and retirement plans.

Over several decades, the Nationwide Finance division had implemented more than 100 legacy applications across dozens of lines of business, functional units, subsidiary companies, and geographic locations. As a result, the finance division grappled with problems common to the large enterprise—inconsistent data, poor visibility, and subpar data governance and reporting capabilities.

Nationwide Finance's implementation of an enterprise data warehouse has unleashed a profound business impact across all aspects of its operations. In production since October 2005, the enterprise data warehouse is at the heart of a revamped finance data infrastructure called FOCUS (Faster, Online, Customer-driven, User-friendly, Streamlined) that has redefined the role and operating model of Nationwide's core finance functions, processes, and systems. Standard data definitions and well-defined data governance are common across the FOCUS platform.

The enterprise data warehouse and MDM platform are enhancing the finance organization's ability to grow value for the various businesses at Nationwide. There is now one authoritative source of consistent, accurate enterprise finance data. All data is clearly defined across the enterprise so each business unit and the three major information systems (ERP general ledger, EPM, and data warehouse) use the same data definitions. Updates are consistent and transparent to all business units, and the data warehouse is supplying the consistent information needed to enable effective decision making across the enterprise. Management has easy access to relevant and timely information and can compare results across the enterprise with confidence and minimal manual intervention. Through FOCUS, Nationwide is reinventing the way finance services are performed and delivered to the enterprise.

CATEGORY

PREDICTIVE ANALYTICS

Pannon GSM

Pannon GSM, a wholly owned subsidiary of Norway's Telenor ASA, holds a share of more than 34 percent in the Hungarian mobile telecom market. Hungary's mobile telecom market has a penetration ratio of 87.5 percent, so Pannon GSM focuses heavily on customer retention and loyalty, churn management, and revenue protection. That means it needs the ability to mine customer and product data to predict behavior and reach each customer on an individual basis.

Pannon had operated with disparate systems and reporting sources that provided slow, inconsistent data; no common business language across departments; and no ability to automate marketing processes such as campaign management. Campaigns had to be done manually, and campaign responses could not be handled by the existing systems. In 2000, Pannon implemented a Teradata warehouse to serve as a single source of consistent, accurate data.

Pannon now has one fully integrated enterprise data warehouse. The data warehouse integrates such information as sales, call detail records, customer data, billing and payment, events, customer interactions, external sources, and network performance. More than 150 business users from across the company access the system.

The data warehouse is the foundation that supports a 360-degree view of the customer as the point of integration for 1 to 10 years of historical data for all customer, channel, and network data. The development of a common language for BI data means that data from disparate sources can be easily integrated into the data warehouse with the assurance that it's one common, accurate set of information. Users now have fast access to consistent, near-real-time data to create complex reports and do sophisticated analyses.

Pannon's data warehouse provided the foundation for implementing advanced data mining and predictive analytics. Those capabilities are delivering business value such as significant churn reduction, exceptional customer response rates through highly targeted campaigns, reduced fraud that saves millions, and much more. These capabilities are keeping Pannon a step ahead of the competition in the saturated Hungarian mobile telecom market.

CATEGORY

RADICAL BI AND DW

AmerisourceBergen Specialty Group

SOLUTION SPONSOR: CONVERSION SERVICES INTERNATIONAL (CSI)

As a key division for AmerisourceBergen Corp. (NYSE: ABC), ABSG works with manufacturers to improve their product launches and expand their markets, ensuring that provider organizations receive the specialty products they need and give physicians the resources for their practices. There are eight divisions within the Specialty Group, and they have combined to lift the company to phenomenal heights. Two words have become the mantra for ABSG employees in 2006: “knowledge driven.” In an industry where competition can win business based on price alone, ABSG is betting that being knowledge driven will separate them from competitors. All indicators show that the strategy is working.

Just as the heart pumps blood to all the vital parts of the body, the ABSG enterprise business intelligence team provides critical data to all systems in the organization. For example, the Rebate Analysis and Pricing Program (RAPP) allows ABSG to launch pinpoint attacks on competitors’ weak spots. A Web site enables customers to log in and see their existing contracts, the price for each item, and even enables customers to maximize rebates. For the CRM system, the EBI team developed near-real-time data feeds to support key alert functionality.

ABSG ships 90 percent of its products the same day its customers place their orders. To manage this, the EBI team provides a dashboard at each of the ABSG distribution centers. A near-real-time dashboard projected on a screen in the warehouse provides visibility into the progress of delivered orders as well as the individual performances of the warehouse workforce. In fact, the lowest-performing worker increased his productivity to become the best worker once the dashboard was displayed.

At less than a terabyte of data and a single source, the ABSG environment may appear to be simple. But the environment has been simplified by design so that the team can continue to rapidly provide incremental value. The ABSG EBI team has measured its success not on the technology behind the solution, but rather the end results of what it has offered to the business.

CATEGORY

UNSTRUCTURED DATA INTEGRATED WITH BI/DW

IBM

International Business Machines Corporation (IBM) manufactures and sells computer hardware, software, and services. With over 325,000 employees worldwide and revenues of \$91 billion (figures from 2005), IBM is the largest information technology company in the world.

In late 1997, IBM’s sales and distribution organization launched a business intelligence initiative under the project name of EDGE (Enabling Decisions for Global Execution). The EDGE system consolidates and feeds information regarding sales leads, order load, financials, and fulfillment into a consolidated picture of key business performance metrics on EDGE dashboards for senior IBM executives and managers. These elements of IBM business activity provide a snapshot of key indicators of the sales pipeline, order progression, revenue attainment, customer service, and delivery, and are used throughout IBM’s management system. EDGE dashboards make these indicators viewable by business unit, by product, and by geography. In other words, EDGE dashboards provide concise, consolidated views of a wide variety of data—including unstructured and semi-structured data.

Before EDGE dashboards were established, IBM’s leadership team was often confronted with different versions of reality when reviewing their business performance. Precious time was frequently spent debating whose facts were correct rather than addressing fundamental business issues such as potential revenue shortfalls and competitive situations.

Today, EDGE dashboards help IBM executives identify shortfalls in the sales pipeline and projected revenue early in the quarter, sometimes before the quarter officially begins—when early action can be taken to address them. By understanding the root causes of systemic issues, executives can put programs, incentives, or customer offerings in place soon enough to impact the current quarter results.

Modifications and enhancements to the EDGE dashboards can be made in a matter of minutes or hours, providing the ability to respond quickly to executives’ needs. Continual, steady deployment of new functions and new content has allowed EDGE to deliver small, frequent victories—and maintain the strong executive support that has been key to its success over the last nine years.



arcplan, Inc.
1055 Westlakes Drive
Suite 175
Berwyn, PA 19312
888.272.0040; 610.902.0688
Fax: 610.902.0689
info@arcplan.com
www.arcplan.com

arcplan, a leading vendor of process-centric business intelligence (BI) software, enables companies to create BI analytics to improve business processes that span multiple IT systems. Customers—such as DaimlerChrysler, CUNA Mutual, Office Depot, and Bayer—rapidly develop customized and interactive BI applications for planning and budgeting, dashboards and scorecards, and operational reporting. arcplan process-centric BI analytics are deployed cost-effectively by leveraging existing data in real time. arcplan has more than 2,500 customers and 300,000 users worldwide. arcplan delivers its solutions through a direct sales force and more than 130 partners in over 30 countries. For more information, visit www.arcplan.com.



Business Objects
North American Corporate Head-
quarters
3030 Orchard Parkway
San Jose, CA 95134
800.527.0580, 408.953.6000
Fax: 408.953.6001
www.businessobjects.com/com-
pany/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 platform, BusinessObjects™ XI, offers the industry's most advanced and complete platform for reporting, query and analysis, performance management, and enterprise information management, including data integration, data quality, and metadata management. 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.



DataFlux Corporation,
a SAS Company
940 NW Cary Parkway, Suite 201
Cary, NC 27513
877.846.FLUX; 919.447.3000
Fax: 888.769.FLUX; 919.447.3100
sales@dataflux.com
www.dataflux.com

DataFlux provides end-to-end data quality integration solutions that enable companies to effectively analyze, improve, and control their data, transforming data into valuable corporate information. DataFlux—a wholly owned subsidiary of SAS—enhances the effectiveness of data-driven initiatives across the enterprise, including master data management (MDM), customer data integration (CDI), and product information management (PIM). The DataFlux Data Quality Integration Solution provides a single platform encompassing data profiling, data quality, data integration, data enrichment, and data monitoring—along with methodologies and rules via service-oriented architecture (SOA) for creating consistent, accurate, and reliable data. To learn more about DataFlux, please visit www.dataflux.com.



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

DATAlegro offers data warehouse appliances that lead the industry in price and performance. Using DATAlegro's patent-pending technology, companies can run faster and more complex queries on their data in order to improve their business intelligence. DATAlegro's data warehouse appliances are powered by Intel's multi-core processors, and use MPP architecture along with Direct Data Streaming™ and RAIDW™. Together, these advanced technologies enable query speeds that are 10–100 times faster than traditional data warehouse technology—at a fraction of the cost.

DATAlegro's high performance appliances offer table scanning speeds of up to 6 terabytes per minute and cost from \$130,000 for a complete system. Whether you have 1 terabyte of data or 300, DATAlegro's appliances have set a new standard in data warehouse price/performance.



ESRI
380 New York St.
Redlands, CA 92373
909.793.2853
Fax: 909.307.3046
gisbi@esri.com
www.esri.com

With annual sales of more than \$610 million, ESRI has been the world leader in the geographic information system (GIS) software industry for more than 30 years. ESRI offers innovative solutions that will help you analyze your data more thoroughly, and visualize and present information more clearly. Working with location information present in your data, ESRI's GIS software and solutions give you the power to solve problems you encounter every day. Public and private organizations in over 200 countries are using ESRI GIS software to make better decisions.



FAST
117 Kendrick Street
Needham, MA 02494
781.304.2400
Fax: 781.304.2410
info@fastsearch.com
www.fastsearch.com

FAST creates the real-time search and filtering solutions that are behind the scenes at the world's best-known companies with the most demanding search problems. FAST's flexible and scalable integrated technology platform elevates the search capabilities of enterprise customers and connects people to the relevant information they seek, regardless of medium.



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.



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
703.848.8600
703.848.8610 fax
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 more than 15 years of industry experience, thousands of customer successes, and a reputation for innovation and leadership, MicroStrategy is the clear choice for your business intelligence investment. More information about MicroStrategy is available at www.microstrategy.com.



Netezza Corporation
200 Crossing Boulevard
Framingham, MA 01702
508.665.6800
Fax: 508.665.6811
salesinfo@netezza.com
www.netezza.com

Netezza, the global data warehouse appliance market leader, enables enterprises to make all of their data actionable—quickly, simply, and affordably. The Netezza Performance Server family of products delivers breakthrough performance, unmatched ease of deployment and operation, and innovative flexibility and scalability at a fraction of the cost of traditional data warehouse solutions. By architecturally integrating database, server, and storage within a single appliance, the NPS system delivers 10 to 100 times the performance at half the cost of existing systems. Customers who have recognized the benefits of Netezza's approach include Ahold, Amazon.com, Cingular Wireless, CNET Networks, Debenhams, Orange UK, and The TJX Companies.



Oracle
500 Oracle Parkway
Redwood Shores, CA 94065
1.800.ORACLE1
www.oracle.com

Oracle's business is information—how to manage it, use it, share it, protect it. The world's largest enterprise software company, Oracle is the only vendor to offer solutions for every tier of your business—database, middleware, business intelligence, business applications, and collaboration. With Oracle, you get information that helps you measure results, improve business processes, and communicate a single truth to your constituents.



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 server, 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: 201.930.8282
info@syncsort.com
www.syncsort.com

Syncsort Incorporated is a leading developer of high-performance storage management and data warehousing software. For over 35 years, Syncsort has built a reputation for superior product performance and reliable technical support. An independent market research firm named Syncsort one of the top Data Warehouse 100 Vendors seven years in a row. Over 90 percent of the *Fortune* 100 companies are Syncsort customers, and Syncsort's products are used in more than 50 countries to back up and protect data in distributed environments, speed data warehouse processing, and improve performance in applications with high data volumes.



XLCubed, Ltd.
(United Kingdom)
27 Market Street
Maidenhead SL6 8AA
+44 1628 763 222
Fax: +44 1628 763 250

(United States)
8001 Irvine Center Drive
4th Floor
Irvine, CA 92618
949.754.3126
Fax: 949.754.4010

www.xlcubed.com
info@xlcubed.com

XLCubed provides the leading direct-to-Excel query, analysis, and reporting environment for Microsoft Analysis Services & Applix TM1, putting real OLAP capability, reporting, analytics, dashboards, and KPIs into the familiar environment of Excel. Our software enables companies to rapidly deploy analytic applications and enterprise reporting solutions by leveraging their installed infrastructure of MS SQL Server, MS Dynamics AX, MS Office, and the Internet. Delivering business solutions across the globe, our customers include Bank One, ExxonMobil, ING Bank, Dell, BP, Cargill, TDC, Grant Thornton, and Microsoft.

About TDWI

TDWI, a division of 1105 Media, is the premier provider of in-depth, high-quality education and research in the business intelligence and data warehousing industry. Starting in 1995 with a single conference, TDWI is now a comprehensive resource for industry information and professional development opportunities. TDWI sponsors and promotes quarterly World Conferences, regional seminars, onsite courses, a worldwide Membership program, business intelligence certification, resourceful publications, industry news, an in-depth research program, and a comprehensive Web site (www.tdwi.org).



MEMBERSHIP

Through TDWI Membership, business intelligence and data warehousing professionals learn about the latest trends in the industry while enjoying a unique opportunity to learn, network, share ideas, and respond as a collective whole to the challenges and opportunities in the industry.

TDWI Membership includes more than 5,000 Members who are business and information technology professionals from *Fortune* 1000 corporations, consulting organizations, and governments in 45 countries. TDWI offers special Membership packages for corporate Team Members and students.

Membership 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
- Consulting calls (Team Memberships only)
- TDWI MemberConnect, an online networking forum
- Access to password-protected, Members-only online content, including TDWI research, publications, and archives
- Discounts on custom research (Team Memberships only)
- Special discounts on TDWI conferences, seminars, CBIP exams, and more

■ www.tdwi.org/membership

WORLD CONFERENCES

TDWI World Conferences provide a unique opportunity to learn from world-class instructors, participate in one-on-one sessions with industry gurus, peruse hype-free exhibits, and network with peers. Each six-day conference features a wide range of content that can help business intelligence and data warehousing professionals deploy and harness business intelligence on an enterprisewide scale.

Conference Highlights:

- More than 50 full-day and half-day courses
- Broad range of course offerings covering fundamentals to advanced topics
- Industry-leading CBIP certification program
- Objective, vendor-neutral coverage of the latest tools and technologies

■ www.tdwi.org/conferences

SEMINAR SERIES

TDWI Seminars offer a broad range of courses focused on the skills and techniques at the heart of successful business intelligence and data warehousing implementations. The small class sizes and unique format of TDWI Seminars provide a high-impact learning experience with significant student-teacher interactivity. TDWI Seminars are offered at locations throughout the United States and Canada.

■ www.tdwi.org/seminars



ONSITE COURSES

TDWI Onsite brings TDWI courses to customer sites and offers training for all experience levels. Everyone involved gains a common knowledge base and learns in support of the same corporate objectives. Training can be tailored to meet specific business needs and can incorporate organization-specific information.

Benefits of Bringing TDWI Education Onsite:

- Discuss your organization's issues openly with TDWI's world-class instructors
- Get TDWI education tailored to your company's needs and objectives
- Skip airfare, lodging, and travel time by training at your workplace

■ www.tdwi.org/onsite

CERTIFIED BUSINESS INTELLIGENCE PROFESSIONAL (CBIP)

Convey your experience, knowledge, and expertise with a credential respected by employers and colleagues alike. CBIP is an exam-based certification program that tests industry knowledge, skills, and experience within five areas of specialization—providing the most meaningful and credible certification available in the industry.

■ www.cbipro.com

WEBINAR SERIES

TDWI Webinars deliver unbiased information on pertinent issues in the business intelligence and data warehousing industry. Each live Webinar is roughly one hour in length and includes an interactive question-and-answer session following the presentation.

To streamline the registration process, we've added a new multi-event signup form that allows you to quickly access our upcoming events, as well as our extensive library of archived Webinars.

■ www.tdwi.org/education/Webinars

WHITE PAPER LIBRARY

Visit TDWI's recently launched White Paper Library, which features a wide variety of industry white papers from TDWI's software and consulting partners.

White Paper Categories:

- Administration and Operations
- Analytic Apps and Development Tools
- Business Analytics
- Business Intelligence Services
- Data Integration
- Data Warehouse Design
- Information Delivery
- Infrastructure

■ www.tdwi.org/WP

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.



Connect. Share. Learn.



Enhance and Accelerate Your BI and DW Career

The business intelligence and data warehousing industry is evolving rapidly. And in a challenging and ever-changing environment, maintaining your competitive edge is vital. TDWI provides a comprehensive selection of cutting-edge industry news and information, research, online resources, and peer networking opportunities developed exclusively for its Members. TDWI Membership can give you the information you need to advance your professional development and make your business intelligence and data warehousing projects a success.

TDWI Member Benefits:

- Quarterly *Business Intelligence Journal*
- Annual *TDWI Salary, Roles, and Responsibilities Report*
- Annual *BI Benchmark Report* **NEW!**
- Quarterly *Ten Mistakes to Avoid* series
- Biweekly *TDWI FlashPoint* e-newsletter
- Quarterly Best Practices Research Reports
- Quarterly Technology Market Reports **NEW!**
- Consulting calls **NEW!** (Team Memberships only)
- TDWI MemberConnect, an online networking forum **NEW!**
- Access to password-protected, Members-only online content, including TDWI research, publications, and archives
- Discounts on custom research **NEW!** (Team Memberships only)
- Special discounts on TDWI conferences, seminars, CBIP exams, and more

For more information, please contact:
Erica Shaver, Business Development Manager, TDWI
206.246.5059 ext. 136, eshaver@tdwi.org

www.tdwi.org/membership



What Works

BEST PRACTICES IN BUSINESS INTELLIGENCE
AND DATA WAREHOUSING

Volume 22

1201 Monster Road SW
Suite 250
Renton, WA 98057

425.277.9126
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/ww22whitepapers.aspx>

Analytic Apps and Development Tools

1. Build Better Cubes: Real-Life Advice on Building Analysis Services Cubes
2. Enterprise Information Management: Tentative Past, Promising Future
3. Mitigate Business Intelligence Project Risks With Rule-Based Audits and Proof-of-Concepts

Business Analytics

4. Defining Business Metrics that Matter—Improving Business Results with High Performance BI
5. The Next Generation of Business Intelligence: Operational BI
6. The Requirements of Process-Centric BI: Meeting the Challenges of the Next Generation of BI

Data Integration

7. Automated Data Cleansing Using Relevant Matching
8. Data Governance: Why It Matters to the Business
9. ETPL—Extract, Transform, Predict and Load
10. Five Steps to More Valuable Enterprise Data
11. Getting Started with Master Data Management
12. GIS and Business Intelligence: The Geographic Advantage
13. How to Launch a Data Governance Program: Practical Guidelines for Technology Implementers
14. Integrating Business Reporting & Analysis for Better Business Decisions Every Day

Infrastructure

15. Data Warehouse Technologies: An Architectural Comparison

Data Warehouse Design

16. Joins and Aggregates Are Critical for Achieving Faster Data Retrieval in a Dimensional Data Warehouse