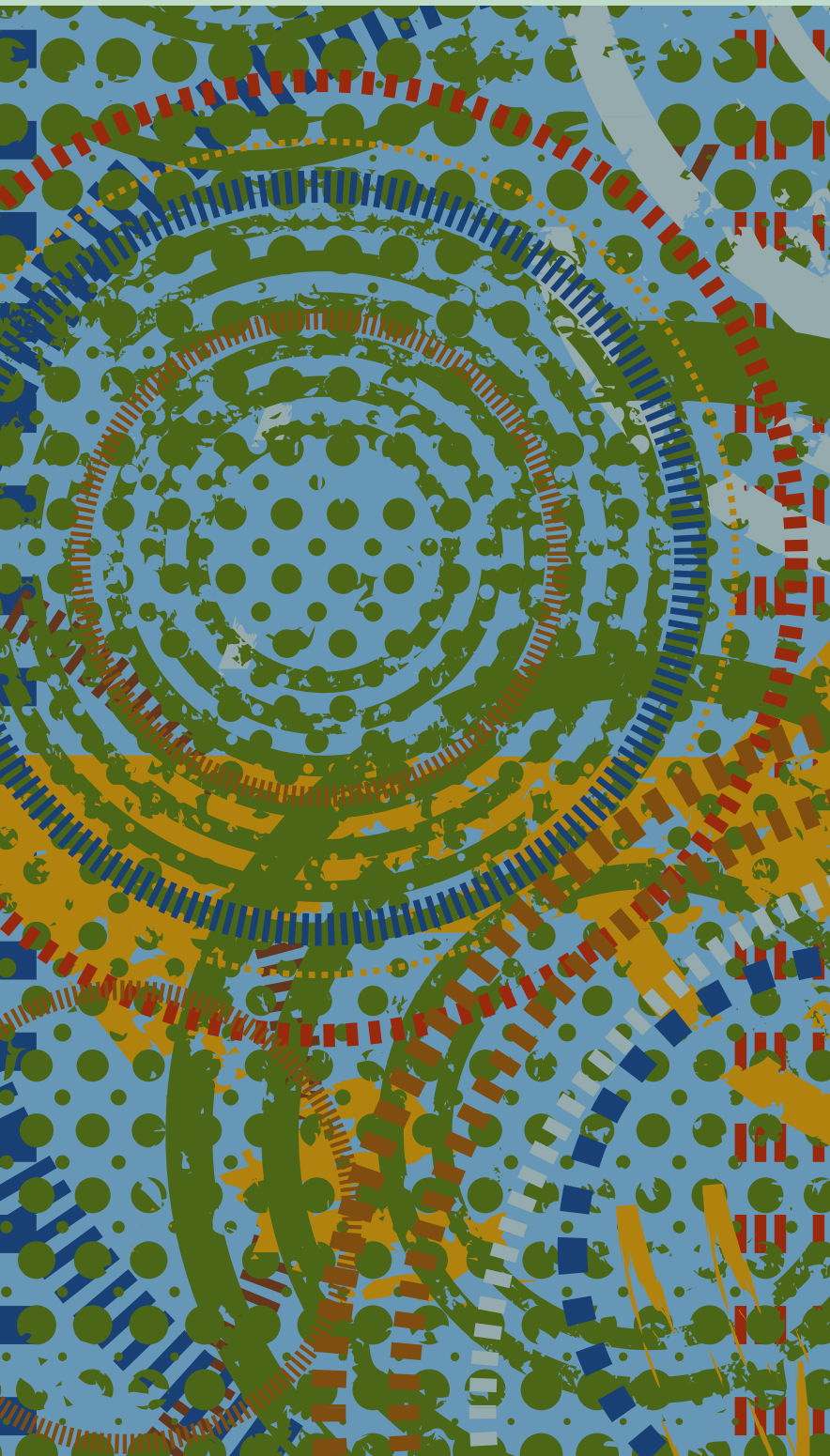


What Works

IN DATA INTEGRATION



**POWERFUL CASE STUDIES, LESSONS LEARNED,
AND Q&A FOCUSING ON:**

- Data Integration
- Customer Data Integration
- Master Data Management
- Data Quality
- Special Section: Data Warehouse Appliances

FEATURE

Operational Data Integration

Philip Russom, TDWI Research

A look at the hottest and fastest-growing practice in data integration today.

page 2

TDWI RESEARCH

BI Search and Text Analytics

page 38

Predictive Analytics

page 49

Letter from the Editorial Director

TDWI is proud to offer you our new, topically focused *What Works in Data Integration*. This collection of customer success stories and expert perspectives provides a resource for better understanding the tools, technologies, and methods that are central to data integration. To help you make your way through the many powerful case studies and "lessons from the experts" articles, we have arranged them into specific categories: general data integration, customer data integration (CDI), master data management (MDM), data quality, and a special section on data warehousing appliances.

So what's inside?

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 data integration technologies and solutions for industries ranging from telecommunications to stock trading.

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 data integration best practices and trends.

Q&A WITH THE EXPERTS

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

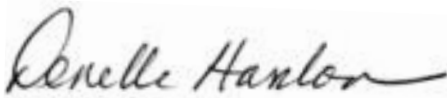
FEATURE ARTICLE

Our feature article comes from Philip Russom of TDWI Research. In "Operational Data Integration," he takes a look at the hottest and fastest-growing practices in data integration today.

TDWI RESEARCH

There's more from TDWI Research. *What Works* includes excerpts from TDWI's recent best practices reports: Search-Enabled BI, TDWI's latest report from Philip Russom, and Predictive Analytics, from Wayne Eckerson.

We hope you enjoy this collection of case studies, best practices, and expert insight that are all focused on data integration and DW appliances. 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 in Data Integration*

TDWI

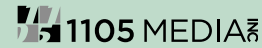
dhanlon@tdwi.org

What Works
IN DATA INTEGRATION



www.tdwi.org

General Manager	Rich Zbylut
Director of Research	Wayne Eckerson
Director of Marketing	Michelle Johnson
Art Director	Deirdre Hoffman
Editorial Director	Denelle Hanlon
Managing Editor	Jennifer Agee
Graphic Designer	Bill Grimmer



President & CEO	Neal Vitale
CFO	Richard Vitale
Sr. VP, Human Resources	Michael J. Valenti
VP, Financial Planning & Analysis	William H. Burgin
VP, Finance & Administration	Christopher M. Coates
VP, Audience Marketing & Web Operations	Abraham M. Langer
VP, Information Technology	Erik A. Lindgren
VP, Print & Online Production	Mary Ann Paniccia
Chairman of the Board	Jeffrey S. Klein

Reaching the staff: Editors can be reached via e-mail, telephone, or mail. To e-mail a member of the staff, use the person's first initial followed by their full last name @1105media.com.

Renton office: 425.277.9126 Fax: 425.687.2842
1201 Monster Rd SW, Ste. 250
Renton, WA 98057

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

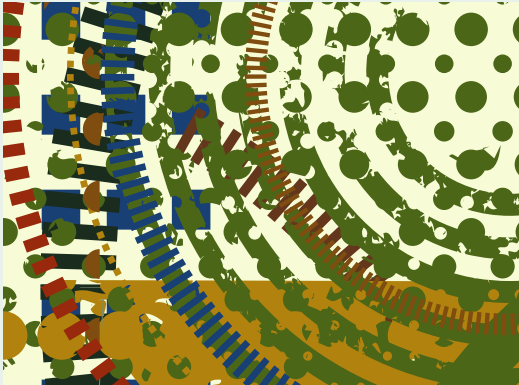
Reprints: For editorial and advertising reprints of 100 copies or more, and digital (Web-based) reprints, contact PARS International at 1105reprints@parsintl.com or 212.221.9595, or visit www.magreprints.com/quickquote.asp

List rentals: This publication's subscriber list, as well as other lists from 1105 Media, Inc., is available for rental. For more information, please contact our list manager, Merit Direct. Phone: 914.368.1000; e-mail: 1105media@meritdirect.com; Web: www.meritdirect.com

© 2007 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 In Data Integration*, 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.

Table of Contents



FEATURE

- 2 Operational Data Integration
Philip Russom, Senior Manager, TDWI Research

An introduction to one of the hottest and fastest-growing practices in data integration today.

DATA INTEGRATION CASE STUDIES AND LESSONS FROM THE EXPERTS

- 6 Data Integration Defined

DATA INTEGRATION

- 7 Real-Time Intelligence Helps Traders Stay Ahead of a Bull Market
- 8 Corporate Express: Using Business Intelligence to Improve Efficiencies and Enhance Customer Value
- 9 Exploring the Next Generation of Information Dashboards
- 10 Iron Mountain Conquers Data
- 11 Data Integration: One Size Does Not Fit All
- 12 S&H Solutions: A Competitive Edge in Retail Analytics
- 13 Powering Integration and Analytics When the Data Is the Business
- 14 Effective Long-Lived Data Management
- 16 Full-Service Bank Gets it Done Faster with the Right Tool
- 17 Alleviating Data Integration Headaches

CUSTOMER DATA INTEGRATION

- 19 Pitney Bowes Improves Total Cost of Ownership
- 20 Worldwide Leader in Software Achieves Accurate, Complete Views of B2B Organizational Hierarchies
- 21 Customer Data Integration: Global Business Visibility with Organizational Hierarchy Management
- 22 Embarq Improves CRM While Increasing Employee Efficiency and Productivity
- 23 Moving Toward A Master Data Management Solution

MASTER DATA MANAGEMENT

- 30 MedSolutions: Improving Healthcare through Information Access and Streamlined Processes
- 32 Leverage a Disciplined, Standard Approach to Achieve “One-Source” Truth
- 33 Master Data Management: Extracting Value from Your Most Important Intangible Asset
- 34 LexisNexis: Enhancing Sales Campaign Support and Improving Data Quality
- 35 Master Data Management: Myth of the Mega-Vendor Platform
- 36 Gaining New Insights with Master Data Management
- 37 Master Data Management and Business Performance Management

DATA QUALITY

- 42 Emerson Process Management Unifies Account Data for a Stronger Revenue Stream
- 43 Avoid the Pitfalls of Poor Data Quality
- 44 Intellidyn Uses Data Integration to Manage Industry-Leading Consumer Information
- 45 Data Integration Drives Healthy Customer Relationships
- 46 M&S Money Meets the Compliance Deadline for Basel II through an Enterprisewide Data Quality Initiative
- 48 Data Quality Monitoring: The Basis for Ongoing Information Quality Management

SPECIAL SECTION: DATA WAREHOUSE APPLIANCES

- 53 Data Warehouse Appliance Combines Unconstrained Analytics with Energy Efficiency for UK’s Leading Mobile Carrier
- 54 Data Warehouse Appliances Help Tackle the Data Center Power Crisis
- 55 HP-Oracle Reference Configurations for Data Warehousing
- 56 Aquascutum Innovates in Retail
- 57 Investing in Data Warehouse Query Performance
- 58 TEOCO Increases Performance and Customer Satisfaction
- 59 Appliances—Data Mart or Enterprise Data Warehouse?

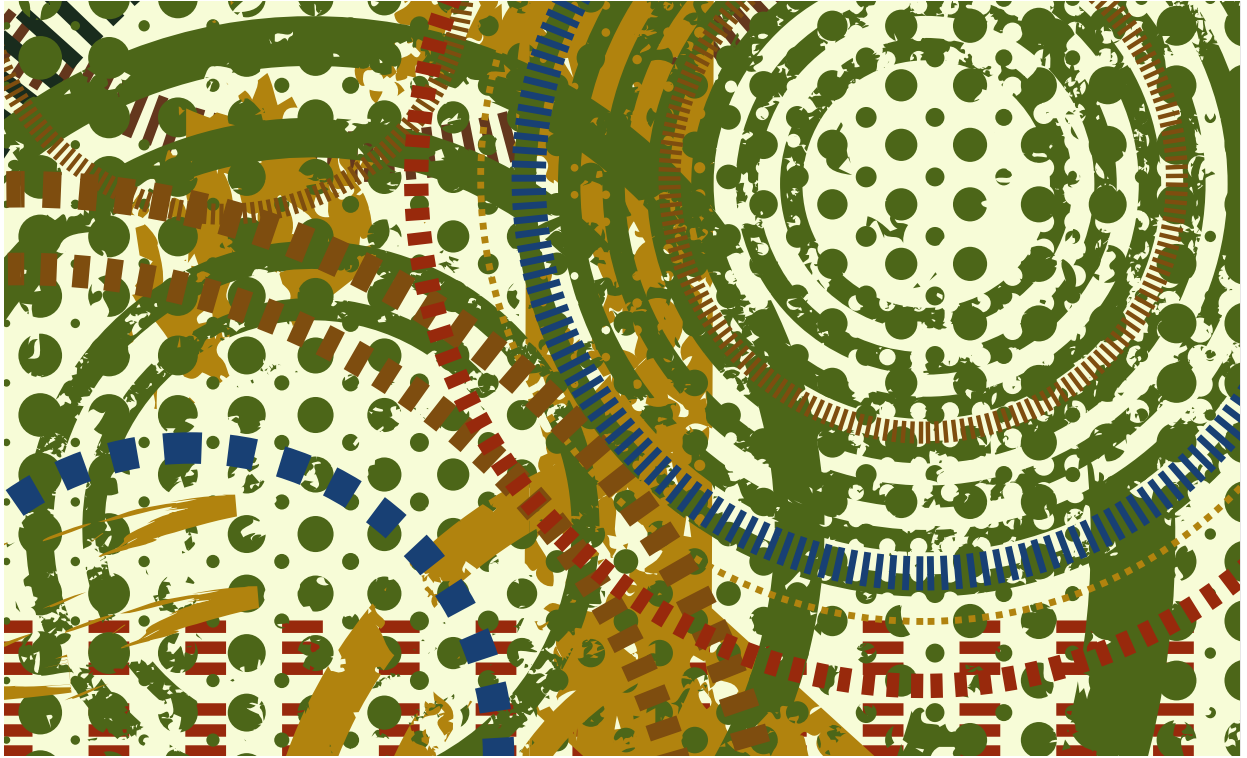
24 Q&A WITH THE EXPERTS

TDWI RESEARCH: BEST PRACTICES REPORTS

- 38 BI Search and Text Analytics
- 49 Predictive Analytics

RESOURCES FOR BI & DW PROFESSIONALS

- 60 Solution Providers
- 67 About TDWI
- 68 TDWI Partner Members



Operational Data Integration

By Philip Russom, Senior Manager, TDWI Research

This article introduces one of the hottest and fastest-growing practices in data integration today: operational data integration. Before we look at different manifestations of operational data integration—like database consolidations, migrations, upgrades, and so on—let’s distinguish it from analytic data integration and explain why you should care.

Distinguishing Enterprise, Analytic, and Operational Data Integration Practices

Enterprise data integration—defined as every imaginable use of data integration—divides into two broad practice categories (see Table 1):

- **Analytic data integration** is applied most often to data warehousing (DW) and business intelligence (BI). It’s also applied (less often) to initiatives like customer data integration (CDI) or master data management (MDM).
- **Operational data integration** manifests as implementations, projects, or initiatives commonly described as the consolidation,

collocation, migration, upgrade, or synchronization of operational databases. This article will look at each of these in detail later.

As a rule of thumb, the database targets distinguish the two practices of analytic data integration and operational data integration. For example, if data integration feeds an analytic database like a data warehouse or mart, then it’s most likely an analytic practice. If it feeds or creates a database supporting an operational application, then it’s probably operational data integration. Of course, data integration tools and techniques enable both practices, and some enterprise initiatives—like operational business intelligence and master data management—straddle the fence to bring the two practices together.

As a quick aside, let’s remember that data integration uses a variety of techniques and tool types, including enterprise application integration (EAI); enterprise information integration (EII); extract, transform, and load (ETL); replication; and miscellaneous utilities.

ENTERPRISE DATA INTEGRATION

Analytic Data Integration	Operational Data Integration
<p>Initiatives commonly supported:</p> <ul style="list-style-type: none"> Data warehousing (DW) Business intelligence (BI) <p>Initiatives less commonly supported:</p> <ul style="list-style-type: none"> Customer data integration (CDI) Master data management (MDM) 	<p>Initiatives commonly supported:</p> <ul style="list-style-type: none"> Database consolidation Database collocation Database migration Database upgrade Database synchronization

Table 1. A taxonomy of data integration practices.

IT professionals implement these techniques with vendor tools, hand coding, or functions within database management systems. It's true that some techniques and tools are closely associated with particular initiatives, like ETL with data warehousing or replication with database synchronization. Yet, all the techniques and tool types under the broad rubric of data integration operate similarly: they copy data from a source, merge data coming from multiple sources, and alter the resulting data model to fit the target system it will be loaded into. Because of the similar operations, industrious users can apply just about any tool (or combination of tools) to any data integration implementation, initiative, or project.

You should care about the distinction between analytic and operational data integration practices because the two have different technical requirements and organizational support:

- Technical requirements.** Analytic data integration usually involves hefty transformational processing to create time series and other dimensional data structures required of a data warehouse. Analytic data integration also supports dozens of data sources and targets coordinated via a hub-and-spoke architecture. By comparison, operational data integration is simple, involving light data model transformations between relatively few data sources and targets (sometimes only one source and one target) connected via point-to-point interfaces. These are the main reasons why ETL and EII—with their unique transformational capabilities and hub architectures—are the preferred techniques for analytic data integration, as well as why the simple interfaces with light transformations found in replication, EAI, and hand-coded practices are sufficient for most implementations of operational data integration.
- Organizational support.** Analytic data integration is usually staffed by the data warehouse team and funded by its sponsor, whereas operational data integration projects are often staffed by a data management or applications group within IT with sponsorship from line-of-business managers and others associated with the applications. As the number of operational data integration projects has increased in recent years, IT manage-

ment and others have drawn data integration specialists from the data warehouse team to perform operational work. This is a problem, because it sets back the goals of data warehousing and goes against the grain of organizational structures and funding. Staffing both camps with data integration specialists is a common solution, despite the resulting redundancy of personnel. To avoid such redundancy, some corporations create a data integration competency center, which staffs both analytic and operational data integration practices as a single resource via shared services.

Problems that Operational Data Integration Addresses

Before describing operational data integration implementation types, let's step back and consider why these are necessary.

Redundant data and non-standard databases are the main problem. For example, when data repeats across multiple databases, it's hard to keep the databases synchronized. Likewise, data may reside in a legacy database that is beyond its prime or in a database brand that is simply not the corporate standard. To put it another way, these are problems because they increase IT costs and inhibit unified visibility into business processes. In a related trend, many organizations today try to “do more with less” and centralize both IT and business operations. These situations eschew redundancy and promote standards; thus, they seek solutions via database consolidations, migrations, and so on.

Redundant and non-standard applications are a problem, too.

An implementation sometimes focuses on an operational application that's a legacy needing migration to a more modern brand or an application instance that's redundant and should be consolidated with other instances. Because there's a database in the application's technology stack, some form of operational data integration is required to migrate or consolidate the database.

Implementations of Operational Data Integration

Now that we've defined analytic and operational data integration practices and explained why you should care, let's look at some of the benefits and challenges of common implementations.

Database Consolidation. This is where IT personnel consolidate multiple, similar databases (often with redundant content) into a single one, with a single data model. If a single database isn't a realistic goal, then data is at least consolidated into fewer databases. The most common project is probably the consolidation of redundant CRM or ERP applications and their databases. As another example, many organizations have multiple mid-tier databases containing slightly different views of customer data, and these are ripe for consolidation. On the analytic side of data integration, consolidations are a popular way to reduce the number of redundant data marts.

The upside of database consolidation is that it reduces IT maintenance costs by consolidating data into fewer servers. Furthermore, it increases visibility into business processes by putting “data eggs” in fewer baskets. The downside is that consolidation is extremely intrusive, in that it kills off the consolidated systems. Their owners, sponsors, and users may resist passing control, typically to a central organization, and these people are usually forced to change business processes associated with the consolidated systems.

Database Collocation. Collocation and consolidation are similar. In a database consolidation, diverse data models are merged into a single data model in a single database instance. This is hard work, so sometimes it's faster and easier to collocate databases. Collocation copies multiple databases into a single instance of a database management system (DBMS), typically on a single hardware server, without merging the data models. In other words, one DBMS instance manages multiple databases. While collocation reduces IT maintenance costs, it doesn't solve the information silo problem like database consolidation does.

Database Migration. In a migration, database brand A gets migrated to brand B. Typical examples include migrating a database from a mainframe to a less expensive open system, from a legacy platform to a more easily supported one, or from a non-standard brand to one that's a corporate standard. Migrating to a newer or standard database increases the flow of information among systems by making data access and integration easier. This way, database migration contributes to advanced integration goals, like real-time or on-demand information delivery.

Like consolidation, migration is intrusive, because it kills off the original system, forcing changes to business processes and applications. With database migrations off of legacy platforms (especially the mainframe), you need to check contracts before committing. Sometimes issues of licensing, leasing, depreciation, or amortization can halt the retirement of an old or non-standard system. And beware that database migration is something of a myth; it's more like new development when the target system doesn't exist and must be designed and built from scratch.

Database Upgrade. This is where version X of a database brand is upgraded to version X+1 of the same brand. A database upgrade

can be a discrete project or an interim step within larger operational data integration implementations. For instance, it might make sense to upgrade all databases to the same release before consolidating them. And upgrading a legacy or non-standard database to the most recent version first can make it easier to migrate. Sometimes multiple upgrades are required, say from version 6 to 7, then from version 7 to 8.

But is a database upgrade really data integration? Yes, it can be. For instance, when customizing a packaged application alters its database's data model, upgrading the application requires operational data integration to upgrade the database. Furthermore, when a new version of a DBMS changes how it stores data, a database upgrade can require transformational processing to get the data into a model optimized for the new version. Finally, when IT decides to improve a data model in tandem with an upgrade, some form of operational data integration is required.

Database Synchronization. The types of operational data integration we've seen so far—database consolidations, collocations, migrations, and upgrades—move whole databases in one-off projects that rarely require that integration infrastructure be left in place. Database synchronization is different, in that it leaves databases in place and exchanges data among them, which requires a permanent integration infrastructure for daily data feeds. The downside of synchronization is the initial cost and maintenance cost of the infrastructure. The upside, however, is that synchronization is non-invasive because it leaves database investments and the business processes that depend on them intact. In fact, when it's not possible to consolidate or migrate databases, managing redundant data usually involves keeping it synchronized across related databases.

High-end implementations of database synchronization include transformational processing and many-to-many data movement, and so ETL is a popular choice. But the reality is that most implementations move data only from one database to one other with little or no transformation. For these, replication and even EAI are suitable techniques.

All other implementation types considered here are inherently one-way and latent. But database synchronization can address unique requirements for two-way data movement or real-time information delivery. Two-way data synchronization has the added burden of resolving conflicting data values, which is best done with high-end replication or ETL tools, or possibly an EAI tool. Real-time data synchronization is usually done with replication, or sometimes with EAI.

Operational Data Integration Implementations Compared

So far, we've seen that database consolidations, migrations, and upgrades can address the problems of redundant and non-standard databases, thereby reducing IT administrative costs and increasing business visibility into operational data. When these

IMPLEMENTATION TYPES	BENEFITS		CHALLENGES				
	Improvement (to data & biz)	Immediate Fix	Initial Cost	Maintenance Cost	Development Time	Intrusion	Overall Risk
Migration	●	●	●	○	●	●	●
Consolidation	●	●	●	○	●	●	●
Collocation	○	●	○	●	○	●	○
Upgrade	○	●	○	●	○	●	○
Synchronization	●	○	●	●	●	○	●

Table 2. Operational data integration implementation types compared.

Key: ○ = Low ● = Medium ● = High

aren't possible, database synchronization and collocation can alleviate some problems in a less invasive way.

That's the good news, and it's great to have options. The bad news is that—given the long list of options—it's sometimes difficult to decide which type of operational data integration implementation (or combination) is most appropriate for a particular situation.

Table 2 assesses the benefits and challenges of the operational data integration implementation types discussed in this article. The assumption is that selecting one or more implementation types involves a trade-off between benefits and challenges. For example:

- Database migration and consolidation are great immediate fixes that can add value to data, despite high risk because of their highly intrusive nature.
- Collocation and upgrades are low-risk because they're cheap, fast, and relatively non-intrusive. But they add little or no visible value to data and business processes.
- Database synchronization improves data content tremendously in a non-intrusive way, but suffers high initial costs (especially when you first put in integration infrastructure) and administrative costs over time (because you must maintain the infrastructure).

You needn't agree with all the assessments in Table 2, because you can adapt it to reflect your own assessments and those of your colleagues. You might also expand the table to include other implementation attributes that matter to you, like which data integration techniques (ETL, EII, EAI, etc.) are best suited to each implementation type and how applicable each is in your organization (perhaps based on preexisting tools and expertise). Regardless of how you adapt it, a table based on consensus-driven assessments can give

the decision-making process more consistency and efficiency. In addition, it will serve as an established vocabulary to get communication and collaboration started.

Recommendations

Segregate operational data integration from analytic data integration. The two involve different types of implementations (or projects, if you will), although both can be done with similar techniques and tool types, like ETL, replication, EAI, and so on. And the two usually have separate sponsorship, funding, and staffing. A possible exception to this recommendation is to bring the two practices together in a data integration competency center.

Address redundant or non-standard databases. This is done with various implementations of operational data integration, namely database consolidation, collocation, migration, upgrade, and synchronization. These projects are worth doing, because reducing the number of redundant databases and moving data to standard databases reduces IT administrative costs and increases visibility into business processes.

Expect to apply multiple implementations of operational data integration. For example, database collocations and upgrades are typical prerequisites for database migrations and consolidations—or vice versa, in some cases. And some databases may need synchronization after they've been migrated or consolidated.

Realize that database consolidations and migrations aren't always possible. After all, these are highly invasive. But database collocation and synchronization get similar results less invasively. Apply your alteration of Table 2 to selecting an operational data integration implementation type or a combination of types. ●

Data Integration Defined

To help you make your way through the many powerful case studies and “lessons from the experts” articles in *What Works in Data Integration*, we have arranged them into specific categories: general data integration, customer data integration (CDI), master data management (MDM), data quality, and a special section on data warehousing appliances. What do these terms mean, and how do they apply to your organization?

DATA INTEGRATION

page 7–18

Fundamentally, data warehousing is an exercise in data integration. A data warehouse attempts to re-integrate data for analytic purposes that organizations have maintained in multiple, heterogeneous systems. Pulling together and reconciling dispersed data is a difficult task. Data needs to be accessed and extracted, moved and loaded, validated and cleaned, and standardized and transformed. Data integration tools support all these processes and make it possible to execute the rules created by developers in the design phase of data warehousing.

CUSTOMER DATA INTEGRATION

page 19–23

Customer data integration (CDI) is the set of processes, controls, automation, and skills necessary to standardize and integrate customer data originating from different sources. More than ever, companies need to be able to recognize different individuals—often encompassed by the term “party”—across diverse business lines and touch points. CDI allows them to uniquely identify parties by reconciling data across different siloed systems. This means coalescing diverse data to identify a single customer across products, purchase transactions, sales channels, salespeople, subsidiaries, and geographies. This implies generation of the so-called “golden record”—the authoritative and certified profile about an individual customer. ... It’s apt that the processing platform that reconciles and integrates data is known as a “hub.”

MASTER DATA MANAGEMENT

page 30–37

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

DATA QUALITY

page 42–48

Data quality is a complex concept that encompasses many data management techniques and business-quality practices, applied repeatedly over time as the state of quality evolves, to achieve levels of quality that vary per data type and seldom aspire to perfection. The criteria for measuring quality vary according to the type of data, its use, business requirements, what’s possible technologically, and how tolerant the technology and the business are to defective, incomplete, or non-standard data. Data management techniques are usually implemented through homegrown or vendor-built software solutions, the point being that software provides automation for measuring quality and adjusting data.

DATA WAREHOUSE APPLIANCES

page 53–59

A strict definition of data warehouse appliance is: “Server hardware and database software built specifically to be a data warehouse platform.” A looser definition allows appliances to be hardware and software designed for any purpose, though bundled and pre-integrated for data warehousing. In a February 2007 TDWI Technology Survey, roughly half of respondents chose the strict definition, a quarter the loose one. However, the focus of data warehouse appliances is shifting from proprietary to commodity hardware, as well as more generally from hardware to software components. In fact, some of the newer data warehouse appliance vendors openly describe their products as software-based accelerators, not hardware boxes. When added to a user organization’s existing BI technology stack (or another vendor’s appliance), these accelerate BI development, and—once in place—they accelerate query performance.

References

- Dyché, Jill, and Evan Levy [2006]. *Customer Data Integration: Reaching a Single Version of the Truth*, John Wiley & Sons.
- Russom, Philip [2006]. *Master Data Management: Consensus-Driven Data Definitions for Cross-Application Consistency*, TDWI Best Practices Report, October.
- [2006]. *Taking Data Quality to the Enterprise through Data Governance*, TDWI Best Practices Report, March.

TDWI Marketplace: Data Integration
www.tdwi.org/MarketPlace/category.aspx?catId=37

CASE STUDY

Real-Time Intelligence Helps Traders Stay Ahead of a Bull Market

Commentary by Sharief Zaman
IT Director, SwiftTrade

When regulatory changes revolutionized the business of day trading in April 2001, SwiftTrade made a smooth transition to proprietary trading, a new type of enterprise in which a contract trader works on a single corporate account. This trading model reduces the individual account costs that were the demise of daytrading practices. In order to manage the rapid growth that ensued, we sought a business intelligence and data integration solution to manage a rapidly expanding network of traders at dozens of offices around the world. Our business has grown from 20 branches two years ago to more than 70 branches today, with an additional 125 planned in the near future.

All of this growth motivated SwiftTrade to build a more effective system for delivering information to every member of our organization. In an industry where every second of the workday is precious, we knew our traders would insist on nothing short of accurate, real-time information—delivered dynamically when they needed it.

To create a new data integration layer and BI architecture, SwiftTrade purchased Information Builders WebFOCUS Business Intelligence Platform and iWay Universal Adaptor Suite. These software products allowed us to build a new information-delivery environment that works efficiently with vast amounts of operational data, from invoicing to branch-wide trading statistics, and it makes information available to the right people at the right time. Trading information that was formerly stored in proprietary back-office applications is now accessible via a reporting database that is continually refreshed with current data.

The integration architecture is straightforward. First, we transferred operational data from a Sybase database on a Sun Solaris

platform to a Microsoft SQL Server database on an Intel platform, which consists of Dell blade servers running the Microsoft Windows 2003 operating system. Then we established a security architecture that governs each user's ability to access certain layers of information based on their roles, positions, and security clearances.

.....

With the data integration solution now in place, trading information that was formerly stored in proprietary, back-office applications is accessible via a reporting database that is continually refreshed with current data.

.....

Trading Up

Once the basic integration and security infrastructure was in place, we created a reporting portal using the WebFOCUS Business Intelligence Dashboard that features about a dozen standard reports, including branch enumeration figures, daily and historical profit and loss reports, settlement reports, account reconciliation reports, invoicing reports and payment reports, and a comprehensive bird's-eye view of the entire company from the head office perspective. These reports are delivered to the dashboard via WebFOCUS Reporting Server, which accesses Microsoft SQL Server data via the iWay SQL Server Adapter.

The new integration layer is fast, flexible, and scalable. Each iWay adapter consists of a communications interface, a SQL translator, and a database interface that translates American National Standards Institute (ANSI) SQL into native application programming

interface (API) calls. The iWay Universal Adapter suite includes more than 300 adapters to connect almost any information system, so we have plenty of flexibility should we decide to adopt a different computing platform or database management system in the future.

With the data integration solution now in place, trading information that was formerly stored in proprietary, back-office applications is accessible via a reporting database that is continually refreshed with current data. We can isolate the information that is pertinent to just one individual trader, or to one branch, or to the entire firm, depending on what we're looking for.

In addition to the nightly data migration procedures, we used the integration technology to create a trigger-based system that automatically updates key reports whenever designated revenue and trading milestones are reached. Real-time reports are ideal, because traders don't like waiting around or wasting time looking for the information they need. Now we can put that information right in front of them, and use the system to recognize outstanding traders. It's a great motivational tool. ●

[Click here](#) to request more information about Information Builders.

CASE STUDY

Corporate Express: Using Business Intelligence to Improve Efficiencies and Enhance Customer Value

Commentary by Matt Schwartz

Director of Business Analysis, Corporate Express

Corporate Express, Inc., a Buhrmann company, is one of the world's largest business-to-business suppliers of essential office, facilities, furniture, and computer products and services, with 2005 sales of approximately \$4.6 billion in North America and \$7.3 billion worldwide. Corporate Express's product offerings include office and computer supplies, imaging and computer graphics supplies, office furniture, document and print management, desktop software, promotional products, and other similar products. Corporate Express has operations in Australia, Austria, Belgium, Canada, France, Germany, Hungary, Ireland, Italy, Luxembourg, the Netherlands, New Zealand, Poland, Spain, Sweden, the United Kingdom, and the U.S. Through exclusive strategic partnerships, Corporate Express can deliver products to China, Finland, Hong Kong, Malaysia, Mexico, Norway, Portugal, Singapore, Slovenia, Switzerland, and Taiwan. Corporate Express's North American operations have more than 200 facilities, including 38 distribution centers, and the company employs 10,775 people. The company's Web site is www.CorporateExpress.com.

Challenge: Creating Business Reviews for Customers

Every quarter, the Corporate Express sales organization creates about 15,000 business reviews for customers using Microsoft PowerPoint. These presentations allow Corporate Express customers to examine their purchasing patterns by product category and evaluate ways in which they can improve their purchasing efficiency. This used to be

a very time-consuming process. In the past, there were 28 versions of the PowerPoint template (instead of a single, centralized template); data had to be pulled manually from multiple reports and from ad hoc queries; and it took up to six hours to prepare each business review.

Solution

Matt Schwartz, director of business analysis at Corporate Express, learned about MicroStrategy Office when he attended the MicroStrategy World 2005 user conference in Las Vegas. MicroStrategy Office is a plugin to Microsoft Excel, PowerPoint, and Word, which gives Microsoft users easy access to any MicroStrategy report. MicroStrategy Office users can run MicroStrategy reports and documents directly from their Microsoft Office applications and refresh the reports with a single click. Since MicroStrategy Office is built on the MicroStrategy platform, all MicroStrategy platform security features, report formatting, and prompts work as well for MicroStrategy Office users as they do for MicroStrategy Web and Desktop users.

After returning from the conference, Schwartz worked with Matthew Cryer, a business intelligence analyst/developer at Corporate Express, to modify an existing PowerPoint business review template to embed MicroStrategy reports. After approximately two weeks of development, Schwartz showed the new, 25-slide PowerPoint document to some sales executives to get their buy-in and feedback.



"The sales executives were really excited about the new tool," Schwartz said. "The time savings and customer value were immediately obvious."

Impressive Time Savings

Using the new, centralized template, salespeople who need to create business reviews can simply hit the refresh button on the MicroStrategy toolbar and answer two prompts to identify the customer and date ranges of interest. At that point, MicroStrategy Office takes over and automatically refreshes the presentation with the latest MicroStrategy reports. The typical refresh and formatting time is less than one hour.

"The time savings that we will realize are truly significant," Schwartz said. "We now have an approach that creates professional business reviews quickly and efficiently. Additionally, using a single template for our reviews means that all of our customers will receive consistent information."

Corporate Express originally piloted the MicroStrategy Office-enabled process with 7 of 28 U.S. divisions, and later rolled the process out to all 28 U.S. divisions.

"The PowerPoint Business Review Tool has been fully deployed to all geographic markets in the U.S. with great adoption," Schwartz said. "MicroStrategy has been instrumental in achieving greater efficiencies and improving our business performance." ●

LESSON FROM THE EXPERTS

Exploring the Next Generation of Information Dashboards

By **Wende Cover**

Director of Strategic Marketing, MicroStrategy

Organizations rely on business intelligence software to tap into their databases and uncover analytically based insights to improve decision making and business performance. Through effective reporting, analysis, and monitoring of organizational data, companies can find new revenue opportunities, improve operational efficiencies, reallocate resources, and reduce costs.

Dashboards are useful tools for monitoring the status of key performance indicators (KPIs) and business metrics. The recent introduction of dynamic dashboards has elevated the functionality and visualization capabilities of dashboards, enabling users to interact with their business data in compelling new ways. Let's explore some of the common misperceptions about dashboards and how this new caliber of dashboards can help users manage their business areas with greater insight, efficiency, and transparency.

Dashboards Are for Executives Only

Historically, dashboards were provided by the IT department to give company executives a daily, weekly, or monthly snapshot of the business. While dashboards are still an excellent tool for executives, they can now be deployed across the entire enterprise and provide useful tools for all levels of business users. In fact, business users can even build their own personal dashboards using intuitive, drag-and-drop design techniques and simplified dashboard templates directly from their Web browser, without requiring assistance from IT.

Dashboards Offer Limited Ways to View Data

Typical dashboards offer standard formatting and data presentation options such as graphs, grids, tables, and text fields. The

latest dashboards offer a wide range of information exploration options, including integration with Adobe Flash technology, making them highly responsive to users. Business users can instantly flip through many different perspectives of corporate performance to quickly identify problems and diagnose root causes without ever leaving the dashboard. Today's dynamic dashboards also offer time-series animations that can convey more meaning from every set of data. Viewing time-series data as graphical "movies" gives business users deeper insight into business trends. Users can rewind and fast-forward through the time series and even pause the movie to drill down for more detail on areas of interest.

Dashboards Provide a Snapshot of Information

While dashboards serve as an excellent tool to monitor the health of an organization or specific business area, dashboards now offer the ability to drill into data, sort and segment data, and interact with data in exciting new ways to gain enhanced insight for decision making. One dashboard can replace as many as 15 different reports, allowing decision makers to see many sets of data in parallel, without changing context or focus.

Dashboards are Accessible Only when Connected to the Network

Have you heard about portable, self-contained dashboards? The Flash rendering found in dynamic dashboards allows business users to create self-contained dashboards that include full interactivity, visualization functionality, and data content. Users can interact with their dashboards on airplanes, present their dashboards in meetings, and analyze corporate performance at any location, regardless of network availability.



Figure 1. Dashboards are useful for monitoring KPIs and business metrics.

Dashboards are Only Useful at the Departmental Level

The new breed of dashboards provides complete transparency across the enterprise by aggregating information from many databases and presenting the data in an intuitive and visual manner on cross-discipline, information-rich dashboards. Dashboards can be used to align organizations through common goals and metrics. And, with automatic personalization and security, a single dashboard can be deployed throughout an enterprise, with each user seeing only their slice of the enterprise and the data that they are allowed to see.

Today, business users have the ability to create expressive dashboards with a highly responsive and easy-to-use interface that makes even the most advanced reporting a welcome experience. If you haven't explored the latest in dashboard technology, you will be amazed at the rich functionality available from dynamic enterprise dashboards. ●

For a free white paper on this topic, [download](#) "MicroStrategy Dynamic Enterprise Dashboards," or [click here](#) for more information about MicroStrategy.

CASE STUDY

Iron Mountain Conquers Data

Commentary by Dave Weldon

Vice President of Technology, Iron Mountain

Customer Profile

Iron Mountain stores and protects digital information, physical records, and artifacts for more than 90,000 customers worldwide. Specializing in offsite data storage, disaster recovery, comprehensive records management, and business continuity services, Iron Mountain's annual revenue tops \$2.2 billion.

Challenge

Changing business models, new regulatory requirements, and rapid growth through acquisitions posed data access and reporting challenges for Iron Mountain. "We were consolidating 11 different billing systems to try to understand our customers and what they were buying," says Dave Weldon, vice president of technology at Iron Mountain.

With so many billing systems, it was difficult to get a consolidated view of customers. Further, Iron Mountain tapped into demographic data from Dun & Bradstreet (D&B) to enrich its understanding of customers and prospects—adding another layer to the data complexity. Finally, Iron Mountain aimed to sustain a growth rate of 10 to 15 percent per year, which meant the company needed to be able to access and understand their data in order to identify new customers, as well as sell more to existing ones.

"Adding D&B information to the data warehouse told us a lot more about our customers, their locations, what industries they are in, and how big they are," says Weldon. "Unfortunately, we weren't able to get that information into the hands of our business community. We had a very large database and only about 10 people—mostly marketing folks—that used it."

Turnaround time was also an issue. The sprawling complexity of data sources and sheer volume meant IT analysts took up to two months to deliver hundreds of reports to the business-user community.

"This cycle took anywhere from a day to 60 days, and it was happening at a rate of about 700 to 800 report requests a month," says Weldon. The company needed faster and more accurate reporting, as well as the ability to provide tailored reports drawn from the 180 databases.

"Getting a worldwide snapshot of a customer and what they were buying used to take two months, and the results were not very accurate. Now it takes two minutes—and it contains three years of historical buying trends."

Dave Weldon, Iron Mountain

The Result

Iron Mountain looked to Business Objects to help them integrate, manage, and report on data stored in systems across the organization. Solutions implemented included BusinessObjects Data Integrator, BusinessObjects Data Federator, BusinessObjects RapidMarts, BusinessObjects Enterprise XI, BusinessObjects Web Intelligence, InfoView, and Crystal Reports.

Real-Time Access to Integrated, Trustworthy Data

Iron Mountain uses Data Integrator to consolidate summarized invoice information from their 11 billing systems into their data warehouse, and Data Federator to supplement the historical information with real-time access to production databases.

"Getting a worldwide snapshot of a customer and what they were buying used to take two months, and the results were not very accurate. Now it takes two minutes—and it contains three years of historical buying

trends," says Weldon. "BusinessObjects Data Federator was the only technical solution we could find that did the job quickly and easily."

Increased End-User Adoption and Ease of Use

Iron Mountain implemented Crystal Reports and InfoView to support standard reports from its IT group. "These tools made information accessible to about 85 percent of our business community," says Weldon, "a far cry from the handful of people who were able to use our reports in the past."

Weldon praises Crystal for its prompting, drill-down, and graphic features that make standard reports "easy to use and meaningful," and InfoView for its reporting folders, batch scheduling, file exporting, and e-mail distribution.

Powerful Ad Hoc Reporting

Iron Mountain has trained more than 850 people in the use of standard reports, and another 120 in ad hoc reporting. End users include analysts in marketing and finance, as well as business users from sales and account management. For super users delivering sales, financial, and marketing analysis, Weldon notes, "Web Intelligence is a friendly and powerful ad hoc reporting tool that gives them the level of sophistication and detail they require."

Quality of Service and SLA Performance

Looking ahead, Iron Mountain plans to use Business Objects to manage service-level agreement (SLA) performance. "As we get a more complete picture of our customers and their buying habits, we want to continue filling out that picture by looking at how well we are able to perform to the SLAs we have with our customers," says Weldon.

Access to, and use of, trusted information is truly driving change and business performance at Iron Mountain. ●

LESSON FROM THE EXPERTS

Data Integration: One Size Does Not Fit All

By **Frank Dravis**

Vice President of EIM Strategies, Business Objects

Before your company charges into a data integration initiative, it is important for the project team to carefully identify your requirements and consider what is driving the initiative. Is it discrepancies between results from financial systems and sales? Concerns about the accuracy of information in a compliance report? Lost market share because of a lack of timely information? The end result you are seeking must drive your data integration strategy and design.

The Challenge of Gaining Trusted Information

According to a 2006 survey conducted by Business Objects, only about 10 percent of information workers always have all the information they need to confidently make business decisions.¹ If this is the case, then organizations that successfully provide their employees with access to quality information can truly achieve a competitive advantage.

Data integration is the key to ensuring that trusted information is readily available for end users. Before you select data integration techniques and technologies, a number of important considerations must be weighed to ensure you choose an approach that is best for the current and future needs of your business.

Batch versus Real Time

Historically, the term *data integration* has been synonymous with batch ETL (extract, transform, and load), an approach used to move large data sets from one enterprise application or database to another. This technique is especially valuable in initiatives where data volumes are high, such as data warehousing, systems migrations, and mergers and acquisitions.

But there are drawbacks. The explosion in data volumes is reducing the frequency by which data can be physically integrated, extending update times—for example, from days to weeks. Increasingly, businesses need more agile techniques for integrating data to ensure up-to-the minute information to support global or 24x7 operations and to address the exponential growth in, and management of, corporate data.

Organizations that successfully provide their employees with access to quality information can truly achieve a competitive advantage.

For organizations that depend on immediate information, real-time data integration techniques should be considered. But how fast is fast enough? The “right time” depends on business needs. For example, a company that manages a city rail system must have accurate and immediate information regarding on-time or delayed arrivals and must be able to drill down to individual trains that are experiencing technical problems. Changed data capture (CDC) can be deployed in this scenario to detect data changes in real time and move only the updated information.

Physical versus Virtual Data Integration

Data integration teams must also consider whether data needs to be physically moved or whether a virtual or “in-place” approach to accessing and aggregating data makes more sense. The technique most often applied here is known as enterprise information integration (EII) or data federation. Data federation employs a variety of specialized query techniques to draw the right information from operational systems quickly for immediate analysis without negatively impacting the performance of source systems.

Data federation is ideal for gaining timely access to changing information, and is extremely valuable in industries such as financial services or government agencies, where strict compliance regulations prevent them from replicating certain data.

One Size Does Not Fit All

When was the last time you purchased a “one-size-fits-all” item? Chances are the item didn’t really fit or suit you and ended up at your favorite charity. Don’t make the same mistake with data integration.

Analyze your business requirements carefully to ensure your proposed solution is the right fit. But don’t be afraid to start with one approach and add others as your business needs evolve. Many organizations marry a number of techniques to get exactly what they need. A solid data integration strategy can help you make more effective decisions based on trusted, complete data and transform the way your business works. ●

For a free white paper on this topic, [download](#) “Data Integration: The Key to Effective Decisions,” or [click here](#) for more information about Business Objects.

¹ Based on a recent survey of information workers in the U.S., United Kingdom, France, and Germany, commissioned by Business Objects and conducted by Harris Interactive, June 2006.

CASE STUDY

S&H Solutions: A Competitive Edge in Retail Analytics

Commentary by Sam Morales

Data Services Manager, S&H Solutions

S&H Solutions was looking to expand their hardware capacity to manage significant growth as well as provide more flexible reporting. Via their Sybase IQ environment, S&H Solutions was able to increase end-to-end performance by 500 percent. This environment has given S&H the ability to grow and to undertake new business intelligence initiatives.

As the original creator of customer loyalty, S&H Solutions began over 110 years ago with the nationally recognized customer rewards program S&H Green Stamps. Today, S&H Solutions has revolutionized their retail business offerings by providing sophisticated technological capabilities and knowledge tools that enable retailers to deliver in-store, real-time, customer-specific messaging.

S&H Solutions has recently added an entire new layer of buying pattern analytics to the established loyalty formula. This gives S&H Solutions a leading competitive position in providing retailers with proprietary customer loyalty solutions. Retail partners can now better analyze the precise buying patterns of repeat customers.

The technical challenges are many for S&H Solutions: it must manage integrating data from a variety of sources, all in different formats, run this data through sophisticated data analysis, and deliver meaningful information to its customers—all in a high-volume environment. Sam Morales, data services manager at S&H Solutions, briefly describes the service provided to retail partners: “In short, we provide our partners [the retail store owners] the ability to target their customers [in] real time and on a one-to-one basis. The S&H system interfaces directly to the retailer’s point-of-sale system and provides detailed reports on shopper behavior and promotions. The customers’ unique

loyalty card enables the one-to-one communication and data capture.”

Customer purchase information data is captured in detail and stored in the S&H proprietary data warehouse, regardless of the customers’ membership status. This information can be reviewed historically to provide retailers with information on customer shopping behavior patterns and promotion acceptance. Data collection can also allow for product-type reporting, including inventory management and customer response effectiveness.

S&H Solutions has revolutionized their retail business offerings by providing sophisticated technological capabilities and knowledge tools that enable retailers to deliver in-store, real-time, customer-specific messaging.

The cornerstone of the analytics solution is information—massive amounts of it. The S&H data warehouse stored in Sybase IQ is close to 1.5 terabytes when compressed, which would require at least 3 terabytes if it were uncompressed in a raw format. The main individual item detail table has 4 billion rows, and joins with a 500,000-record Universal Product Code (UPC) table of product lookups. This transaction table also cross-references 10 million customers selected across different combinations of hundreds of individual stores.

This volume of information not only allows global trend analysis; it also provides analytics at the individual product and consumer

level. Based on report results, customers can be individually identified, segmented, and offered behavior-specific messaging and promotions. Each customer is identified with a member number through the S&H system located at the point-of-sale register. The system will notify the cashier in real time with shopper-specific details such as whether the customer is a loyal or high-spending shopper. This real-time connectivity and one-to-one customer communication allows the retail partner to track customers on a continuous basis, and develop high-level customer promotions and messaging to increase shopper frequency and loyalty.

Through high-performance technology and ongoing improvements, S&H Solutions is able to provide competitive customer loyalty solutions to its retail partners to deliver an integrated data set, high-volume consumer intelligence, as well as timely information. S&H Solutions’ solutions for its retail partners’ loyalty programs have proven effective with high-level reports that highlight virtually every aspect of each unique buyer segment. Sybase IQ’s high volume capacity brings a critical richness to the data and an enhanced analytical capability. That information creates consumer loyalty for the stores, and retailer loyalty for S&H Solutions. ●

Powering Integration and Analytics When the Data Is the Business

By **Sumit Kundu**

Director, Product Management, Sybase

Data aggregators are information brokers, collecting industry- or society-wide data to provide value-added services to customers and subscribers over the Web. The services provided by data aggregators go by many names: mortgage risk intelligence, audience measurement services, market research provider, national statistical agencies, online shopping price comparison, and many others. They may be government agencies, independent companies, or an enterprise division.

Complex Challenges

A data aggregator faces the challenge of converting vast stores of data into a product or service, since their operations depend on reselling data and/or related analytics. Some aggregators manage the most complex and demanding data integration and analytics challenges imaginable: capturing millions of data points at the most detailed level from hundreds or thousands of locations every day, only to turn that data around for customer use in very short order. Others face more modest requirements. Typically, all will face one or more of the following challenges:

- Integration of multiple, disparate data sources
- Large numbers of complex, ad hoc queries
- Large data sets, frequently more than a terabyte of data
- Large numbers of concurrent users
- The need for an “active warehouse” of data

Integrating the Data Aggregator’s Environment

Sybase Data Integration Suite addresses the complex data integration needs of data aggregators by combining a modular set of technologies with advanced modeling, development, and management tools. The five core integration technologies provided by the suite are replication, data federation, ETL, real-time events, and search. This comprehensive approach ensures that any data integration requirement, or set of requirements, can be addressed by the suite. Supporting these component technologies is an integrated modeling and metadata layer that leverages the data modeling capability of Sybase PowerDesigner, the industry’s number one data modeling solution. In addition, the suite’s development tool, WorkSpace, provides an integrated development environment (IDE).

Some aggregators manage the most complex and demanding data integration and analytics challenges imaginable: capturing millions of data points at the most detailed level from hundreds or thousands of locations every day.

The Key: Advanced Analytics Capability

Sybase IQ empowers data aggregators to not only overcome their data management challenges, but to thrive in the face of the most demanding environments and business requirements. Sybase IQ’s column-based architecture enables ultra-high performance to support large numbers of ad hoc and

complex queries. Sybase IQ’s data compression algorithms cost-effectively support the very large databases that are typical in data aggregator environments. In addition, Sybase IQ’s multi-node architecture helps data aggregators to simultaneously support a large clientele of concurrent users while allowing continuous data feeds into their data warehouses.

In a benchmark exercise, Sybase IQ was loaded with one trillion rows of data, which is the equivalent of 155 terabytes of raw input data. Sybase IQ used only 55 terabytes to store the 155 terabytes of input data. Conventional databases would require up to one petabyte (1,000 terabytes) of storage for the same amount of input data. In other words, thanks to its unique compression capability, Sybase IQ can reduce data storage requirements by up to 94 percent. Even more impressively, Sybase IQ showed no slowing in query or data loading speed as query submission rates increased five-fold—demonstrating linear scalability. And with customer databases of over 40 terabytes in production, it’s evident that these performance numbers aren’t limited to benchmarks.

In summary, Sybase IQ and Sybase Data Integration Suite provide the capability and performance necessary for data aggregators to integrate complex environments, as well as to recognize and extract an in-depth understanding from large reservoirs of information—making it quickly available to those who need it the most. ●

For a free white paper on this topic, [download](#) “Right-Time Business Intelligence: Optimizing the Business Decision Cycle,” or [click here](#) for more information about Sybase.

Effective Long-Lived Data Management

LESSONS FROM THE OIL & GAS INDUSTRY

By David Archer

Vice President, Petris Technology, Inc

Amy Price

Director, Petris Technology, Inc.

An Introduction to Oil and Gas Information

Throughout the 20th century, technological advances have increased the ability to find and produce oil and gas. Petroleum lies waiting to be produced in reservoirs below the surface of the earth—through deep wells through which oil and gas flows to the surface where it is collected, transported to refineries, and converted into any number of end products that the modern world requires for its current existence.

Since the early 1980s, information technology in the form of complex sets of technical geoscience applications have helped companies discover, define, and develop their oil and gas reservoirs. Because hydrocarbons are “way down there” where we can’t go, industry specialists have come up with a sophisticated array of indirect methods to determine where the hydrocarbons are, how much might be there, how we might drill wells to get to them, and how to extract them. All of these indirect methods rely upon volumes of data that must be properly managed, making this industry one of the most information-intensive in the modern world.

Long-lived Assets with Long-Lived, Large-Volume Data

The productive life of a petroleum reservoir may span many decades, and the reservoir’s data has a lifetime that is at least as long as that of the physical asset. Data accumulation begins at the earliest phases of exploration with volumes of data growing until the property is fully depleted. The data itself is complex to acquire, understand, and interpret. It’s produced by and used in a variety of processes in very dynamic workflows, over long time frames subject to changing asset

ownership, technical/business teams, and technological paradigms.

A major exploration and producing company has many terabytes of data under management—with another terabyte per day of operational data being added to the already overflowing data stores. Not to be neglected are the rooms full of paper documents (and their electronic counterparts).

The productive life of a petroleum reservoir may span many decades, and the reservoir’s data has a lifetime that is at least as long as that of the physical asset.

Data Management Barriers

Oil and gas data is complex, generated and used over a long time frame by many people working across several technical disciplines and is managed by a variety of means which are generally related to how the data was originally captured—without regard to future needs. It is not uncommon to have the same information in several different places—either replicated or used inconsistently. The data itself is in a variety of formats bound to different technical applications and organizations—few of which follow common access methods or use common representations. Attempts to manage all the data in central repositories designed to provide universal access to people and process have proven difficult to create and maintain.

More than 15 years of attempts to standardize industry data models have had limited success; comprehensive data models have

been developed but data tends to remain in the applications that created it. It has proven too costly for existing applications to change their underlying data access mechanisms to be tightly bound to industry data standards. Furthermore, changes in energy and information technologies result in maintenance nightmares for most standardization initiatives—as by their nature standards are resistant to change.

Promising standards efforts in energy (as in other industries) have centered on the definition of XML specifications for information exchange via Web services and service-oriented architectures (SOA). However, these efforts alone are not enough.

Two recent studies of oil and gas information management and integration highlighted some of the barriers to data management in energy. These barriers are not unique to energy. (See Table 1).

A Systems Approach to Data Management

Previous attempts at data management have failed because the problems are systemic; any solution to one aspect will also affect the others. A systems solution—one that considers each of these aspects—provides optimal results.

- **Data should remain in its original repository, while making it visible and accessible** to users of the original and other applications. Underlying applications should not be changed.
- **Use Web-based conventions to search for and transfer needed data sets.** Web-based maps provide a convenient way to locate geo-referenced data (much of the oil and gas data is connected to a

GARTNER	UPSTREAM CIO
Inability to access data	Technical information captive in “silos” or applications
Too many data sources	Large volumes of data
Inconsistency of available sources of information	
Lack of timeliness of information	
Inaccurate / poor quality information	Poor quality data / gaps in data
Lack of integration	Lack of a common integration framework
Standards need to support <ul style="list-style-type: none"> Regulatory reporting Asset M&A Portfolio review Production optimization 	Lack of standards and differences in naming conventions

Table 1. Two recent studies of oil and gas information management and integration highlighted some of the barriers to data management in energy. Sources: Gartner Presentation—Schlumberger Information Solutions Forum, Las Vegas, NV, 2005; Upstream CIO Newsletter, 2006

specific geographic location). Google also has provided a model for how text search can work efficiently, including a way to index and search metadata rather than searching the entire data set. And the “shopping cart” model for transferring data is intuitive to users.

- **Provide the means for new data sources and applications to be added without requiring any change to those already in place or any loss of existing data types.** Applications called adapters connect to each data source and permit the data from that application to be mapped to a model that is the union of all data types connected to the system. This aspect enables the use of standards of virtually any variety. Petris offers the Dynamic Common Model that provides this unique capability.

- **Employ tools and processes to ensure data quality.** When a full collection of data is made visible for the first time, even in companies that have maintained fairly rigorous quality standards the data is generally of very uneven quality, with many “holes” and questionable data sets. Good data management requires good data, and the approach should include the provision of tools and workflows that ensure it stays that way.
- **Reduce the volume of persistent data.** Only the searchable part of the data—the metadata catalog—needs to be stored separately from the data and made available. This catalog should include not only the metadata, but the relationships between the data types. When dealing with large volumes of data, this lightweight search approach is the only

practical way to accomplish the task of identifying needed information.

- **Deploy via the Internet and build on IT standards** such as Web services and SOA. That makes this approach a universally applicable and adaptable to virtually any industry—not only oil and gas. The PetrisWINDS Enterprise’s Web Services module goes beyond this, allowing an application to access any other connected application or data source.
- **Use data management as a way to preserve and transfer knowledge.** With an average energy worker age of around 50, waves of retirements are coupled with “fresh-out” recruiting, leaving few people in between to provide training and mentoring. Good data management allows not only the results but the knowledge (interpretations and supporting documents) to be captured and made accessible. This heritage is invaluable to the life of the asset, while the supporting documents provide a degree of training and insight to younger employees.

A Framework for Innovation

The energy industry has many data management “lessons learned” that are applicable to any industry. We believe that the exciting and powerful advances Petris has made toward managing complex technical information have created an “innovation platform” that includes a Web services interface that enables innovative applications to gain access to the entire collection of data that is connected to the framework. ●

For a free white paper on this topic, download “The Business Value of Integrating and Managing Information,” or [click here](#) for more information about Petris Technology.

CASE STUDY

Full-Service Bank Gets it Done Faster with the Right Tool

Commentary by Lia Szep

Senior Technical Analyst, Syncsort Incorporated

Background

With the right tool, data integration can be quick and painless. The following is an account of one company's search for that right tool. The company, a full-service bank that we will call FSB, operates banking offices in a number of states. In addition to a wide range of traditional banking services, FSB offers a comprehensive array of investment, mortgage, and insurance services, and has a network of loan origination offices for small businesses nationwide.

Challenge

FSB developed a production system to identify and monitor a customer's total relationship with the bank. The system warehouses data primarily to support the lending function of the bank. To populate the data mart that supports this system, data is taken from multiple disparate sources, including mainframe legacy and Oracle systems, converted to a common format (ASCII flat files), and then processed with an ETL tool. The result is that users of the system can view multiple obligations—commercial loans, mortgages, credit cards, deposit accounts, and so on—of an individual customer as a single lending relationship. The system provides debt and deposit totals at both the customer and the lending relationship level, making it possible to create credit presentations, management reporting, or loan portfolios.

The problem FSB faced was in the volume of data. The system had to process 5.5 GB of demographic and account-related data for four affiliate banks.

Solution

The development team at FSB was tasked with evaluating different tools and selecting a solution that could handle this process. The decision came down to three options.

The first option was to convert the text files to Oracle tables, perform a series of joins in Oracle, then extract the data back into text files. The second option was to use Ascential DataStage, the ETL tool they were previously using. The third option was Syncsort's DMExpress. At the time, DMExpress was fairly new, but FSB had used other Syncsort products in the past, so they decided to evaluate their latest software.

Having dismissed the first option as too cumbersome, FSB decided to evaluate DataStage and DMExpress through a proof of concept. The proof of concept involved files of fairly large sizes, as well as complex join and filter criteria. Unable to handle large volumes of data, DataStage would often crash before finishing. When it did complete, it would take three to four times longer than with DMExpress.

“Wherever we find the need to process large amounts of data in a short amount of time, DMExpress is the obvious choice.”

Lead Developer, Full-Service Bank

With DMExpress, FSB was able to process data from four banks—11 files per bank with a total volume of 5.5 GB—run 23–25 DMExpress tasks for each bank, and produce 8 files with a total volume of 1.5 GB. With DMExpress, this process took less than an hour.

FULL-SERVICE BANK PROCESSES DATA MORE THAN 3 TIMES FASTER

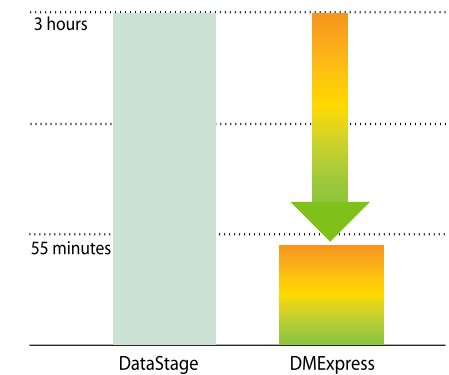


Figure 1. FSB was able to process data from four banks in less than an hour.

Conclusion

After a successful proof of concept, FSB decided there was no other option. “Wherever we find the need to process large amounts of data in a short amount of time, DMExpress is the obvious choice,” the lead developer of FSB commented. When it comes to speed, DMExpress gets it done faster. ●

LESSON FROM THE EXPERTS

Alleviating Data Integration Headaches

By Lia Szep

Senior Technical Analyst, Syncsort Incorporated

When it comes to data integration, IT professionals need more than aspirin to cure their headaches. The challenge of reaping actionable information from unrefined data existing in disparate sources is, to put it mildly, overwhelming. For those charged with integrating and maintaining a company's data, the key to alleviating the headaches of data integration is to identify the challenges ahead of time.

Growing Data Volumes

As businesses thrive and customer bases enlarge, data volumes have skyrocketed. Many large corporations maintain data warehouses teetering on the line between terabytes and petabytes. Take, for instance, one of the most successful retailers in the world: with more than 170 million customers weekly, Wal-Mart was once touted by Tera-data as having the largest database in the world. Forrester Research estimates that the average growth rate of data repositories for large applications is 50 percent each year. Even that may seem like an understatement when you consider Internet-generated transactions. Data warehouses of companies with significant Internet presence can grow by billions of records daily. Mining a large amount of data is daunting enough; add exponential growth rates, and the task seems insurmountable.

Decentralized Information Systems

With expansion, mergers, and acquisitions, most corporations today have more than one location. No matter how well planned business expansion is, a central data hub may no longer be practical. And while decentralized information systems may seem a necessary alternative, multiple systems often translate to major disparities. Data exists in multiple formats, on disparate sources, and is often duplicated elsewhere. Forrester estimates that the latter is the case with 35 percent of all application data. Also, whether it's budget concerns or an unwillingness to change tried-

DIRECT MARKETING COMPANY DRAMATICALLY REDUCES PROCESSING TIME

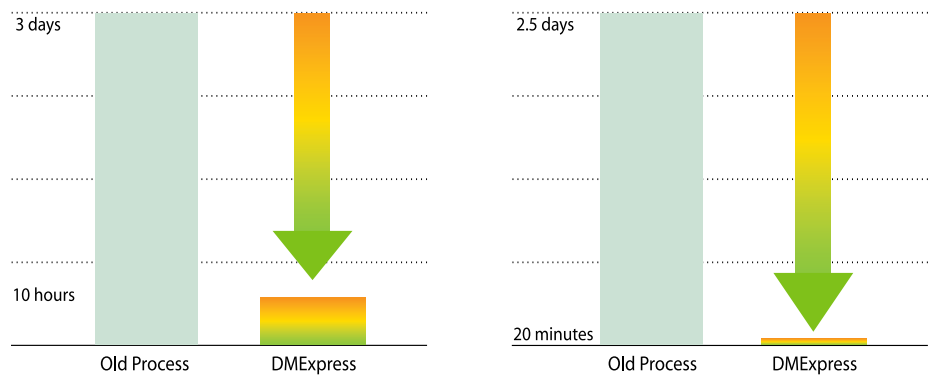


Figure 1. The company can now process data fast enough to run reports on a weekly basis.

and-true procedures, the internal workings of an individual IT department can also play a part in the disparity.

For those charged with integrating and maintaining a company's data, the key to alleviating the headaches of data integration is to identify the challenges ahead of time.

Migrating Legacy Systems

Most IT environments run a number of different platforms. While certain platforms may become less cutting-edge or less proficient with time, the applications that reside on them may still be critical to the company. The best option for maintaining mission-critical processes and increasing the performance of the application is to migrate it to another platform. While this may be necessary, it's no small task. The planning phase alone is cumbersome and time-con-

suming, and the implementation of legacy migration can become a painful game of trial and error.

Demand for Timely Information

A company's success comes in part from its ability to process customer and market data quickly enough to anticipate and respond to changing business trends. With this necessity to analyze mission-critical information comes an incessant demand to get it done faster. For the IT professional, it may seem that processing data faster is never fast enough.

Cost Control

With exponentially expanding data volumes and an increasing demand for faster analysis, the number of applications being developed and supported in most organizations has multiplied. Without a corresponding increase in staff, there are major implications for IT professionals. This can often mean increased workload, less time for individual projects, and more deadlines.

CONTINUED ON PAGE 18

Finding a Solution

There is relief—but you have to find it. This can be the biggest challenge of all. In your search for the perfect solution, always keep in mind the other challenges you will face. If, for example, you are dealing with a vast amount of data on disparate sources, then you'll want to find a solution that runs on multiple platforms and provides support for different sources and targets.

Perhaps the most important objective will be to minimize the run time of an application. No matter what the other aspects of your system might be, unless you can minimize the time it takes to complete a job, you have no chance of meeting management's demand for timely information. People sometimes think that adding hardware will solve performance problems or elapsed time issues. There are a number of reasons why hardware alone won't do it. For starters, run times are almost never linear with data volume growth—when you double the amount of data, you more than double the processing time. Second, doubling the number of CPUs doesn't double the throughput. And finally, adding processors only improves performance if the application is able to make use of the additional processors. In short, hardware upgrades are expensive, and the reality rarely meets the expectation. Since it is impossible to successfully respond to the challenge of exponential data growth with a “hardware-only” strategy, software must be a part of the solution.

The solution must also be scalable. As companies go from gigabytes to terabytes and ultimately petabytes, scalability becomes a necessary tool in managing exponential data growth. Other things to consider include ease of use and vendor reputation. If a product is easy to use, it can help control the cost and time of training staff members. And if the solution is from a reputable vendor, it is more likely to have been proven reliable.

Finally, whatever other options you may be considering, testing the product in your own environment, with your own data, is absolutely the only way to determine a product's performance. Not only must the solution be the best—it must be the best for you.

Real-Life Example

Our customers typically see real value in evaluating our software solution in their own environment. One direct marketing company was able to see significant performance improvement before committing to a solution.

The company's primary focus is building and managing customer databases for *Fortune* 1000 corporations. This provides the necessary framework for organizations to aggressively apply database marketing strategies to their marketing programs. Many of these customer databases house data on nearly every individual in the U.S. With anywhere from 250 to 300 million names, addresses, and other demographic information, extracting demographic data, analytics, profiles, and model scores for processing can be a cumbersome and time-consuming task.

With the tool they were using, it would take three days for this company to run analytics on 240 GB of data. Through a software evaluation in their own environment, the process was completed in less than 10 hours. They were able to assess the performance improvement and determine other processes where the solution could bring them valuable results.

In another project, the company was processing data on the mainframe. The project involved importing data to the mainframe, scheduling the job, running the processing, and outputting the data to a flat file. All of this would occur while other processes were running on the mainframe. Because of this overload, the entire job would take two to three days to complete. Using the high-performance software solution, they were able to completely remove the mainframe from the process and use a much smaller system. The entire job was completed in 20 minutes (see Figure 1). Because of the improved performance, the company can now process data fast enough to run reports on a weekly basis.

Conclusion

Data integration doesn't have to be a major headache. Identifying the challenges ahead of time—and finding a solution that fits your specific needs—is easier than it sounds. Just remember that performance is the bottom line. Minimized elapsed time, combined with scalability and ease of use, adds up to increased performance. Also, in evaluating a solution's performance, be sure you can do so in your own environment. This is the only way to determine the best solution for you. ●

For free white papers on this topic, [download](#) “The Critical Business Need to Reduce Elapsed Time” or “Speed Data Warehouse Processing with High-Performance Joins and Aggregates,” or [click here](#) for more information about Syncsort Incorporated.

CASE STUDY

Pitney Bowes Improves Total Cost of Ownership

Commentary by William Duffy

Data Warehouse Project Manager, Pitney Bowes

OVERVIEW

Industry

High Tech

Oracle Products

- Oracle Business Intelligence Applications
- Oracle Business Intelligence Suite Enterprise Edition
- Oracle Data Warehouse

Key Benefits

- Improved total cost of ownership with comprehensive BI solution
- Enhanced sales productivity through better customer insight
- Increased responsiveness and customer satisfaction
- Improved marketing effectiveness with better segmentation and targeting

The Challenge

Pitney Bowes, a \$5.5 billion company, is the clear market leader for postage meters, with 80 percent market share.

With a global customer base of 2 million, one of the top challenges at Pitney Bowes was making sure they maintain their existing customer base while capitalizing on new opportunities for growth.

This meant that the sales organization needed to better understand their existing customers, to ensure renewals and discover up-sell and cross-sell opportunities. They further required information about the potential of new customers by analyzing the behavior of their existing customers.

With such a large customer base, the company had the challenge of ensuring that their service representatives had all the information required to ensure customer satisfaction—whether it be to repair a piece of equipment or to answer billing inquiries.

The challenge in the marketing organization was creating campaigns based on an understanding of the buying and service behaviors of the customer, such that the right products could be offered to them at the right time.

Pitney Bowes needed a solution that could help them overcome these challenges. Furthermore, they needed a technology that could draw insight from more than 10 different legacy systems.

The company embarked on a vision, called “Power of One,” to provide its customer base with a consistent experience across all business units. At the core of making this vision successful was providing information about customers across all customer-facing associates—four call centers with over 1,250 agents and 1,500 field service repair representatives in North America.

The Solution

Pitney Bowes chose Oracle, as its comprehensive solution provided best-of-breed data warehouse, BI platform, and prebuilt BI application products. Furthermore, its hot-pluggable architecture provided the ability to integrate business intelligence into their business processes, derive insight from their legacy systems, and analyze information from both historical and real-time data sources.

By standardizing its customer-facing associates onto the same Oracle BI solution, Pitney Bowes was able to create a 360-degree view of their customers and provide real-time access to this information to sales, service, and marketing organizations. They also achieved a faster time to value by implementing the Sales Analytics, Service and Call Center Analytics, and Marketing Analytics modules of Oracle BI Applications. Pitney Bowes has standardized on an Oracle data warehouse that can easily scale as their volume of information grows. This solution has helped them improve performance and user satisfaction, and provides an infrastructure to grow their business.

Enhancing Business Processes with Pervasive Insight

Pitney Bowes has integrated Oracle business intelligence solutions into their core business

processes, enabling their customer-facing associates to take actions based on customer insight.

Enhanced Sales Productivity

Oracle Sales Analytics has completely turned the tide in their sales organization. It enables them to analyze where sales agents were spending time, ensure efficient customer follow-up, perform better forecasting, analyze opportunities, and track renewals to close. “Oracle BI provides laser-like visibility into the performance of every sales rep,” said William Duffy, data warehouse project manager. “Sales operations can now pinpoint potential issues right away, which earlier would have taken three weeks.”

Improved Customer Satisfaction

Pitney Bowes receives approximately 30,000 calls a day from its customers. With Oracle Service and Contact Center Analytics, they can now measure the performance of their call centers and effectively manage the productivity of their 1,250 agents, thereby ensuring responsiveness and improving customer satisfaction.

Improved Marketing Effectiveness

Oracle Marketing Analytics helps the company drive important customer retention campaigns, targeting customers with expiring leases. It further helps them segment customers and understand the potential of new customers, so that the right customers are called at the right time to improve overall campaign effectiveness.

Superior TCO

“One of the most important values of Oracle BI is its total cost of ownership,” said William Duffy. “With Oracle BI, we were able to deliver over 400 reports to a very large organization with just one person—now that’s cost effective!” ●

For a free white paper on data quality, [download](#) “Transforming Data into Quality Information,” or [click here](#) for more information about Oracle.

CASE STUDY

Worldwide Leader in Software Achieves Accurate, Complete Views of B2B Organizational Hierarchies

**TRANSACTIONAL HUB DELIVERED IN
LESS THAN SIX MONTHS WITH ACCURATE,
SCALABLE SOLUTION**

Commentary by Jim Cushman

Vice President of Architecture, Initiate Systems, Inc.

Company Overview

A worldwide leader in software, services, and solutions that has achieved success with B2B organizational hierarchies.

The Challenge

This company strived to attain a composite view of all individual and organization data about customers, partners, and suppliers. The organization wanted to define the true value of each customer and identify valuable customers within organizations, even when purchasing through multiple channels. Additionally, this company needed to accurately assign and recognize revenue for accounting and sales commission processing through efficient territory management, ensure accurate license positions, assess customer risk and loyalty instantly and accurately, and comply with regulatory compliance concerns on a global scale. Finally, they wanted to dramatically reduce the internal costs associated with attaining data quality, alignment, and governance.

Solution

This software leader selected Initiate customer data integration (CDI) software for a major project to completely reform its global data infrastructure and better identify individuals and organizations.

Initiate Systems developed a B2B data stewardship application to support the business

requirements. Initiate Hierarchy provides search, retrieval, editing, visual hierarchy navigation and management, and advanced compositing capabilities.

Implemented in less than 12 months, the project aligns data records from five countries and individual data from internal customer management systems and external master reference databases, such as Dun & Bradstreet, Experian, InfoUSA, and others. By building on top of enterprise-class, 64-bit products, such as Microsoft SQL Server 2005 and Windows Server 2003, the project was able to align more than 50 million organizational hierarchy records in under six hours, making them available for real-time consumption with the ability to handle thousands of concurrent transactions with sub-second response times.

“Initiate’s advanced matching technology created a global solution for our CDI needs.”

*Senior Director,
Worldwide Leader in Software*

Benefits

The company has enjoyed many benefits since deploying Initiate Hierarchy.

- **Customer satisfaction**—provides a consistent customer experience at all touch points by improving self-service and enabling complete, accurate, real-time views of all of a customer’s relationships with the company.
- **Financial reporting**—streamlines reporting on unparented and misparented accounts, which allows the company to understand and identify its most valuable customers, more accurately recognize revenue, reduce commission errors, and ratify license positions.
- **Field productivity**—ensures synchronization of data across the entire organization, including customer relationship management (CRM), software licensing, and

financial accounting applications, resulting in greater potential for cross- and up-selling.

- **Privacy and legal compliance**—As the compliance landscape continues to evolve, this company is poised to adapt with it. With organizational hierarchy management, privacy preferences are easier to track and respect.

The Future

This company will soon introduce multi-byte data from additional countries and be poised to include their full global organization base, which will approach 150 million organizations spanning 50 countries and 20 languages. They will also roll out a worldwide data stewardship tool that was developed as a thin-client application. And they will integrate in real time with data standardization and business process management tools and services that offer automatic routing, workflow, and decision making.

Later this year, Initiate Hierarchy will be integrated in real time with large-scale applications such as Siebel and SAP, where it will serve as the global customer master. Integrated applications will leverage Initiate Hierarchy to search and retrieve organizations and hierarchies; update, link/unlink, and merge/unmerge organizations; and much more. This offers a centralized, service-oriented architecture (SOA) for customer data quality and management, and will dramatically increase the accuracy and reliability of the data for a fraction of the cost of attempting to perform the same functions in over 50 unique customer repositories.

Within the next year, this company will integrate its entire global individual customer base, amounting to approximately 500 million records, and will use Initiate Identity Hub software to identify all its unique individual, household, and individual-to-organization relationships. ●

LESSON FROM THE EXPERTS

Customer Data Integration: Global Business Visibility with Organizational Hierarchy Management

By **Richard L. Clements**

Senior Director, Product Marketing, Initiate Systems, Inc.

Are you over-extending credit to your business customers because of a lack of visibility into your relationship with those customers in a hierarchy context?

Do you understand relationships between individuals and organizations so that you can find all relevant potential transactions and grow market share?

Are you missing out on volume discounts with suppliers because you don't realize how much you're buying from one vendor?

The Emergence of Organizational Hierarchy Management

Nonintegrated, conflicting views of business and individual relationships across multiple channels, geographies, applications, and hierarchies result in a disparate view of the customer. Companies have a compelling need to understand how customers and suppliers are organized, because they cannot fully leverage those relationships without understanding them in a hierarchy context. Aligning customers and suppliers to a hierarchy enables global visibility, which is strategic for revenue growth, cost reduction, risk management, and compliance. Enterprise customer data integration (CDI) and organizational hierarchy management provide a proven solution to solve this problem.

Companies Struggle with Existing Approaches

Companies have struggled with the need for organizational hierarchy management for some time, and have tried various approaches to address it. Companies today send data to external organizational data providers for processing, buy organizational data

and a matching engine with which to apply it, or build the organizational hierarchies internally. However, these approaches have offered only partial or limited success.

Using CDI as the Solution

Companies are seeking new ways to solve their hierarchy management challenges. CDI enables companies to:

- Consume multiple different sources of hierarchy information while simultaneously comparing and matching hierarchy structures across sources to create a logical master representation of the hierarchy—a single version of the truth.
- Create hierarchies without size or depth limits through a highly scalable model. CDI defines each organizational hierarchy through the use of metadata, covering the hierarchical definitions themselves as well as taxonomies, key attributes, and indices related to each organizational record. The hierarchy is based on the parent/child relationships between the various members of the hierarchy. This simple concept provides the necessary representation to enable efficient search and navigation of hierarchies.
- Classify individuals against the logical master hierarchy adopted from the external source systems by associating individual customers to nodes in the hierarchy through identity linkages. By leveraging a CDI solution that spans the notion of customer contacts within organizations as well as organizations within hierarchies, companies achieve global customer visibility.
- Query, navigate, and view the entire hierarchy based on a single node. This allows roll-up and drill-down through intuitive, Web-based views and powerful search capabilities. Situations may arise

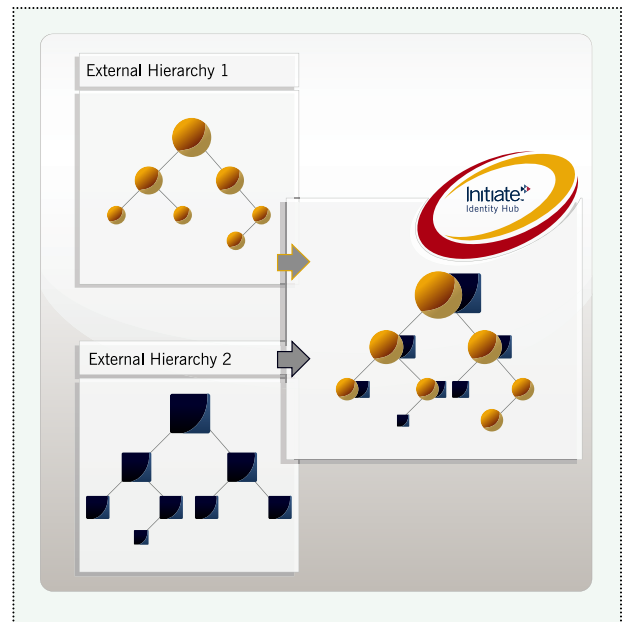


Figure 1. Aligning customers and suppliers to a hierarchy enables global visibility.

where the user needs to modify a hierarchy, altering the linkages. This might be because a conflict has been flagged and the user has elected to resolve the conflict manually, or perhaps a new merger or divestiture has changed the organizational landscape. In addition, CDI makes it possible to split and merge hierarchies, in the case of a change of corporate ownership, for example. And of course, all changes and modifications are logged and audited for future reference.

While a number of technologies understand some of the facets of a single view of the customer and provide functionality with varying degrees of effectiveness, it is only by including the full customer view enabled by CDI and organizational hierarchy management that companies can truly recognize the customer, in the broadest sense of the word, at the point of interaction. ●

For a free white paper on this topic, [download](#) "Addressing the Organizational Hierarchy Challenge," or [click here](#) for more information about Initiate Systems, Inc.

CASE STUDY

Embarq Improves CRM While Increasing Employee Efficiency and Productivity

SOLUTION ACHIEVES SUCCESS WITH NO COSTLY OVERHAUL OF LEGACY SYSTEMS

Commentary by Rob Clucas
 Manager of Database Marketing Analysis,
 Embarq Corporation

Diversity may not be a great idea when it comes to the technology used in managing customer relationships.

Ask the IT staff at Embarq Corporation, a large Midwest telecommunications company. “After a century’s worth of [mergers and acquisitions],” said Rob Clucas at Embarq, “our customer service and sales people had upwards of 40 data sources that they might have to comb through for account planning or to resolve a customer issue.”

Most of the data sources are based on large-scale, heterogeneous, geographically distributed equipment.

One of the strategic challenges for Embarq was to create some means of giving its customer service representatives (CSRs) a comprehensive but unified view of the status of any customer who calls in with an inquiry.

A consolidated view also would support the account planning needs of the sales force. And the solution had to be backed by a highly available, scalable IT infrastructure.

Quick Results

The solution accomplished a virtual integration of customer information without costly and time-consuming infrastructure overhauls and upgrades. It is based on a robust infrastructure consisting of IBM middleware supporting grid computing designed by Nimaya, and implemented through Nimaya’s CustomerGrid.

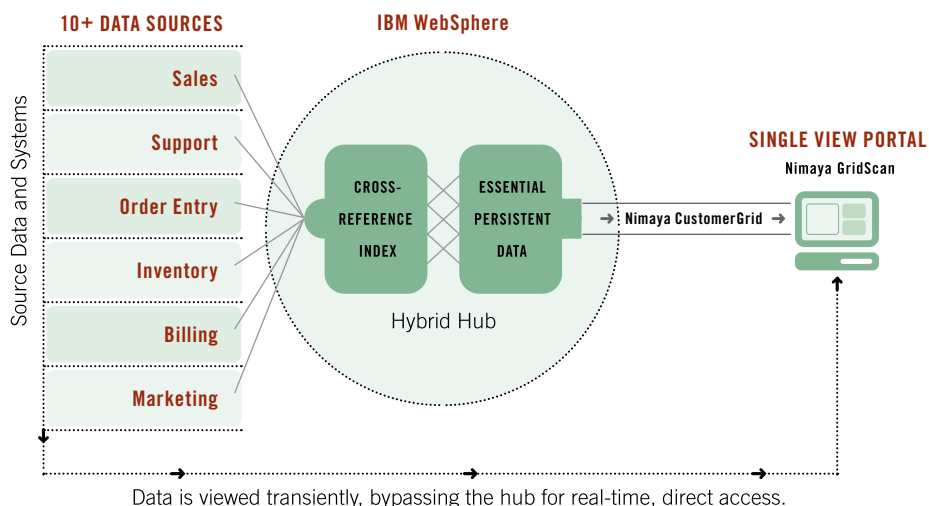


Figure 1. High-level data flow and architecture of the Embarq Solution.

The IBM middleware includes WebSphere Application Server, Portal Enable, and Business Integration Connect-Express.

Rob Clucas indicates that Embarq’s return on investment was “almost instantaneous,” delivering a “highly effective” system. “It’s a much quicker implementation because you’re not having to design and build a hub,” said Clucas. “You don’t have to engage in a huge IT project to get it done. We had prototypes up and working within weeks. Most of the solution elements were out in users’ hands in the first 90 days. There’s no way we could have done that with traditional CDI.”

Initially the solution enabled the 10 most important data sources to be brought together into one screen. These included the basics, such as billing, trouble tickets and dispatch, data-circuit inventory, order entry, account assignments, and marketing data.

“The CSRs love it,” said Clucas, “because they’re responding to questions faster and more easily, and they’re dealing with happy customers. They’re no longer responding to customers with statements like: ‘That’s in a different department—let me transfer you to that group.’ Or, ‘I can’t help you, but I can get you to a person who can.’”

Evidently, the same goes for the sales people. “What we had before was terrible,” said Clucas. “They used to have to keep track of some things manually—literally writing things down on paper as they went from database to database in search of information. If a sales rep needed to see what a customer had, what required changed, or what locations needed a visit, he might have to log on and off half a dozen systems, and he might have to do that 30–40 times a day. Now, through the CustomerGrid, it’s all there on one interface.”

The CSRs and sales force total about 5,000 out of an employee population of 22,000. Such sizable numbers mean that small improvements have significant effects. In Embarq’s case, said Clucas, only nine months after the system rolled out, “we saw an 11 percent drop in the time people were spending on non-sales-related activities.”

Within a year of the initial implementation, Embarq analysis showed that unproductive time had been reduced by 76,000 hours, which is equivalent to about \$2.6 million in annual expense reduction. “Those are recurring savings,” Clucas emphasized. ●

LESSON FROM THE EXPERTS

Moving Toward a Master Data Management Solution

By **Frederick Ferguson**

Associate Partner, IBM Global Business Services

John de Wit

CEO, Nimaya Inc.

Even as network technologies continue to evolve, proliferate, and converge, the telecommunications marketplace has begun to mature. Service providers are shifting their focus from acquisition and usage stimulation to increased account penetration (i.e., wallet share) and improvements in yield (e.g., average revenue per unit, or ARPU). Emerging access methods themselves constitute a proliferation of inbound and outbound sales and marketing channels that must be integrated in near-real-time on a single CRM platform. Ultimately, there is a need to ration and target marketing incentives to customers with actual value improvement potential. To improve and accelerate their business intelligence capabilities, service providers are increasingly integrating master data management into their enterprise data warehouses, data marts, and other analytical infrastructure.

Business intelligence applications consume huge volumes of data such as subscriber profiles, usage records, and billing details. This data is cleansed, integrated, and summarized to provide business decision makers and operational staff with business performance metrics, subscriber scores, and other analytics required to improve overall business performance. Customer-centric metrics (e.g., ARPU, churn rates) typically include the subscriber or subscriber segment as a “dimension.” Other common dimensions include the product category, service area or market, interconnecting service provider, sales channel, and time period. The identifiers that form the domain of each dimension often vary across their (and their business partners’) operational support systems (OSS) and business support systems (BSS) applications. In addition, the relationships between dimensional data (e.g., subscribers in a

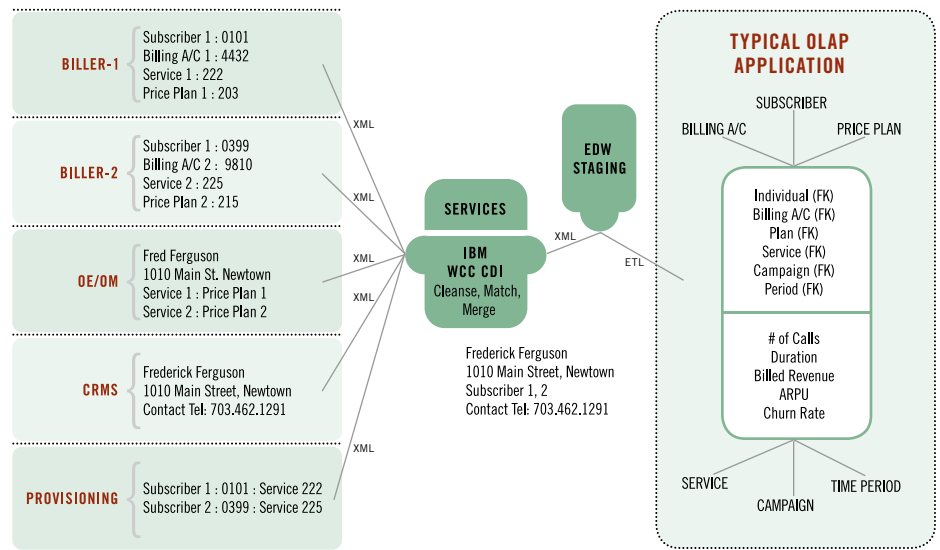


Figure 1. Integrated BI, SOA, and master data management.

household or products in a category) tend to change slowly over time. Data warehouses typically standardize dimensional data using match/merge processes in the staging area to achieve conformity in analytical reporting.

Simultaneously, service providers are deploying master data management (MDM) solutions to integrate, cleanse, persist, and serve dimensional data from across the enterprise for other OSS/BSS applications as part of an enterprise service-oriented architecture. (See Figure 1.) In particular, service providers are deploying customer data integration (CDI) hubs to support a single view of the customer as well as product master hubs to standardize products, their categories, bundles, and components, and their associated features, value-added services, available configurations, and applicable rates. Service providers are beginning to extract still greater return from their investments in master data management solutions by leveraging them as the system-of-record source of dimensional data for the data warehouse, replacing the match/merge processes in the data warehouse staging area. In effect, service providers are moving MDM earlier in the data lifecycle (closer to their operational applications) to better proliferate and reuse data standards across the enterprise.

IBM has invested heavily in development and acquisitions to integrate our suite of

master data management and business intelligence solutions as key components of a next-generation, service-oriented architecture for telecommunications service providers. This integrated approach ensures a single, standard, and reusable view of customer, product, and other master data to improve analytical accuracy, operational efficiency, and business agility. IBM’s MDM and BI solutions leverage multiple IBM and partner solutions for a comprehensive offering. These solutions support robust workflow capabilities by serving master data and analytics in an integrated service-oriented architecture, thus leveraging and enhancing new and existing OSS/BSS investments.

Nimaya’s solution compliments MDM by way of real-time Web services that access disparate data across the enterprise, forming a single view of a customer for proactive alerting, insight, and customer-centricity. This eliminates “swivel-chair” activities and improves productivity among employees. By leveraging MDM/CDI hubs, business rules can be automated, driving workflow activities inter/intra-departmentally, and removing the barriers that inhibit communication among different departments. ●

For a free white paper on this topic, [download](#) “Pathway to Master Data Management,” or [click here](#) for more information about Nimaya Inc.

Q&A with the Experts

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 for each Q&A.

Business Objects

Q We want to enhance our data warehouse with current information stored in our operational systems. How can we do that without degrading source system performance?

A A well-designed data federation tool will allow you to extend your data warehouse with up-to-the-second information from your operational systems while limiting disruption to those systems. It does this through intelligent query techniques, optimized join strategies, specialized algorithms that minimize data transfer, and by leveraging database query optimizers. With these advanced features, data federation can be used to create a virtual data warehouse that provides users with both real-time and historical information without physically moving data or bringing your operational systems to a crawl. In the end, users will have access to trusted, accurate, timely information for better decision making.

ANALYST VIEWPOINT

A federated approach to business intelligence is a sound alternative for organizations needing to track key metrics from source systems with a relatively low cost and rapid implementation, as long as the organization does not require the richer query functionality and long-term historical analysis possible in a traditional data warehouse. Other options to incorporate timely information into a warehouse include real-time trickle feed technology from data integration vendors, as well as changed data capture (CDC), which moves only data updated in source systems since the last batch load to the warehouse. By reducing data volumes, CDC can enable greater load frequency with minimal impact on operational performance.

Collaborative Consulting

Q Does data governance require a budget?

A Absolutely! If you want to accomplish something tangible, you need to obtain management support and an approved budget that weighs costs against benefits. Grassroots data governance usually leads to failure. Identify the goals and objectives of your data governance program, and estimate the resources required to accomplish the goals. Take into account that data governance often requires new software solutions and external guidance. Since effective data governance requires commitment from IT and business users, resources will be diverted from other corporate projects. Make a conscious decision to make the trade-off, and your data governance efforts will flourish.

ANALYST VIEWPOINT

A data governance budget is not only a sound practice to help ensure alignment of resources with business objectives; it can also help prevent “project creep” and unanticipated expenses. Especially for large organizations with a variety of data and metadata management systems, a data governance initiative can take on a life of its own with multiple facets, competing responsibilities, and diluted objectives. A strategic budget, strong executive sponsorship, and cross-enterprise data stewards are key to helping you realize the data governance benefits of common data definitions and closer IT/business alignment across business units. Ideally, a data governance budget will map out total cost of ownership as well as the expected tangible and intangible benefits.

DataFlux

Q What is the role of data quality in master data management (MDM) initiatives?

A MDM is emerging as a way for organizations to deliver a single, unified view of the truth from multiple operational applications. At the core of successful MDM engagements is data quality technology, which helps companies standardize, verify, and correct data across sources. This technology is also adept at matching similar pieces of data in different sources and resolving numerous instances of a customer, product, or asset into a “master record.” Although MDM approaches often focus on the access and movement of data, an MDM approach driven by data quality will lead to better data within the master repository—and a more accurate reflection of the truth.

ANALYST VIEWPOINT

Data quality should be very much a foundational element of an MDM solution. Without data quality in place, organizations run a significant risk of perpetuating the old “garbage in, garbage out” phenomenon across the MDM platform. Close attention to data quality can exponentially increase the value of MDM by cleansing, matching, and standardizing data from operational customer, product, inventory, and other systems. In tandem, data quality and MDM help enable organizations to precisely target customers, eliminate duplicate communications, and improve supply-chain efficiency across regions or globally. Deduplication of data from multiple sources can also help improve MDM operational performance by reducing the number of records to be loaded by a factor of 5, 10, or even more.

DATAlegro, Inc.

Q What are the advantages of using a data warehouse appliance based on a commodity versus a proprietary platform?

A Commodity-based appliances offer significant advantages:

- **Major partnerships.** Commodity players reduce risk by attracting the attention and support of major partners.
- **Rationalized research and development.** Development is focused and advancement is quicker when harnessing the power of major partners to provide hardware reliability and innovation.
- **High availability.** Commodity hardware offers higher reliability and improved mean time between failures (MTBF).
- **Performance and scalability at lower prices.** Commodity players offer higher scalability and performance at lower cost because of strong competition.
- **Easier upgrade path.** Plug and play prevents forklift upgrades.

ANALYST VIEWPOINT

The increasing availability of commodity-based systems has introduced yet another twist in the fast-changing DW appliance market. Many appliances based on commodity hardware (e.g., Intel and AMD microprocessors) and software are competitively priced because vendors can sidestep the dedicated R&D required to develop and evolve proprietary systems, and over time commodity systems generally mature to approach, if not match, the performance of proprietary technology. Organizations in the market for a DW appliance need to weigh price, performance, data volumes, infrastructure integration, and energy efficiency considerations—as well as their long-term road map—in deciding between commodity and proprietary DW appliances.

Hyperion Solutions Corporation

Q What are the key points to consider for an MDM initiative related to performance management and BI?

A MDM projects vary widely, but it’s key to start small and plan for growth from the beginning. First, identify one or two types of master data (dimensions) and hierarchies to tackle. You can start by defining and agreeing on a master data lifecycle, definitions, and attributes, and by determining the system of record and system of entry for changes. Next, identify one or two systems that will interact with the MDM application initially. This means defining the integration method(s) and establishing the required frequency of updates to each system. Finally, ensure business user involvement in the maintenance of master data. This includes securing management sponsorship and participation, establishing policies and business rules for managing changes, and defining change approval levels, internal controls, and reporting.

ANALYST VIEWPOINT

A good rule of thumb when strategizing on an MDM initiative for BI and performance management is to start small but think big. Driven by tactical needs, some organizations apply MDM in isolated areas without a vision for enterprisewide MDM—inviting silo problems and focusing more on reactive MDM rather than proactive MDM. Given differing interpretations of MDM, it’s important to invest time up front in defining what MDM means for the business. Organizations should clearly distinguish between analytic and operational MDM, baseline problems to be addressed, and outline the expected benefits of MDM both immediately and in longer-range pursuit of MDM’s promise of a single view of the business.

HyperRoll

Q What's wrong with data warehouse tuning techniques used today?

A Data warehouse performance tuning techniques are designed to overcome the architectural limitations of the underlying relational database. E.F. Codd, a relational database pioneer, admitted that relational databases need these. Whether you're denormalizing, indexing, partitioning, striping, or creating summary tables or views, you're overcoming existing relational database management system (RDBMS) limitations. These techniques are labeled as tuning, when in fact they're an engineering compromise that trades speed for adaptability, reliability, and cost. With the rapid growth in data and the increasing need for more insightful information balanced with constrained IT budgets, it is time for innovation. Disruptive approaches to high-performance data access are a necessity for success.

ANALYST VIEWPOINT

We all know that RDBMSs were originally built with the relatively minimal data model requirements of transactional applications in mind. The RDBMS vendors have done a good job of extending their database products to enable the much more complex multidimensional data models required of data warehousing. But there's still room for improvement. To fill the void, a few vendors have produced software tools—sometimes called accelerators—that sit atop popular RDBMSs, giving them greater dimensionality in modeling, speed with multidimensional queries, and efficiency in managing dimensional data.

Informatica Corporation

Q What is the role of profiling in data quality?

A The goal of profiling is to identify problems that could prevent effective matching of data. Data quality profiling enables you to answer the following questions about customer data:

- What data fields are suitable for use in the matching processes?
- What standardization/cleansing is required for each data field prior to the matching process?
- What matching rules are likely to be effective? For example, partially incomplete or invalid fields can be used in the matching process, but rules must be formulated to ensure that they are used only when a valid output is present.

ANALYST VIEWPOINT

Organizations that overlook data profiling in a broader data quality initiative do so at their own peril. Though short-term savings might be realized by avoiding both software licensing and personnel costs, skipping the data profiling phase can undermine the integrity of the entire process and mean more time in manually troubleshooting problem spots. When properly executed, data profiling assesses weaknesses and discrepancies in data prior to cleansing and matching and enables reconciliation up front, as well as providing a radar screen for ongoing data monitoring. Especially for high-impact initiatives like customer data integration, data profiling can be the make-or-break element in the project's overall success.

Initiate Systems, Inc.

Q What are the necessary features for accurate and effective customer recognition and data matching?

A **Dual thresholds.** Most systems based on probabilistic algorithms can be tuned to achieve specific false positive and false negative rates. However, look for a system that provides the ability to set multiple thresholds for each search.

Real-time response. Avoid solutions that offload batch processing with no emphasis on performance. Instead, look for a system that can scale to support millions and billions of records for on-demand record lookups.

Adaptability. Businesses concerned with high accuracy should also look for a highly adaptive system—one that adjusts according to the data contained in individual files.

Extensibility. To ensure high accuracy, companies must be able to include search parameters specific to their business or industry.

ANALYST VIEWPOINT

As customer data integration (CDI) systems proliferate, more attention is being paid to the functionality and reach of data cleansing and matching solutions. A checklist feature should be supported for virtually all data types—structured, unstructured, and information from external third parties. The solution should also provide outlier reporting to enable users to monitor its effectiveness and determine the root cause of problems. Because packaged data matching software can be complex, with probabilistic algorithms, heuristics, comparisons, and scorings, organizations may want to consider engaging an independent expert to assess the pros and cons of competing applications and help select the solution that best meets their business requirements.

MicroStrategy

Q How can we efficiently tap into our data to enhance the decision-making process?

A Effective decision making requires a user to view a sequence of interrelated data. With reporting technology, a series of 5 to 10 reports may be needed to make analytically based decisions. This process is time-consuming, both for IT to design the reports and for the businessperson to uncover critical insights.

Dynamic dashboards with relational online analytical processing (ROLAP) functionality offer an advanced approach to decision making. These informational dashboards collapse data from dozens of reports into one dashboard, and provide intuitive and rapid navigation across the data. With ROLAP's "drill-anywhere" capability, users can easily surf the data warehouse to find the data without requiring an explicit report to be designed by IT.

ANALYST VIEWPOINT

A key culprit behind inefficiency in tapping data is user reluctance to embrace sophisticated business intelligence tools. Rather than attempting to force-feed BI down users' throats, many organizations have benefited from lowering the barrier to entry through third-party Microsoft Excel add-ons that enable analysts to continue using a favored tool but with richer ad hoc analysis and reporting functionality. Similarly, dynamic, graphical dashboards with easy-to-read metrics and drill-through, workflow-driven decision making, as well as collaborative training involving both business and IT, can help improve user efficiency. A business requirements analyst can help understand user needs and complaints and tailor solutions accordingly.

Netezza Corporation

Q Data warehouse appliances have been acknowledged as successful at supporting data mart analytics. But can they handle the requirements of enterprise-class data warehousing?

A Increasingly, large organizations are implementing Netezza data warehouse appliances to handle enterprise-class data warehouse applications because of their ability to easily manage large workloads of varying complexity. These systems have proven they can deliver high performance against large, mixed workloads across many business applications and for hundreds of concurrent users. This capability provides organizations with faster, deeper insight into data from multiple departments across the enterprise.

ANALYST VIEWPOINT

Appliances are steadily gaining credibility as scalable platforms for enterprise-class data warehouse systems—one reason why IDC predicts the appliance market to grow from its current estimated \$50–75 million to roughly \$500 million in five years. Continued R&D by DW appliance vendors and the advent of multicore processors are combining to increase appliance scalability to support large data volumes, complex queries, and high numbers of concurrent sessions. Prospective buyers should conduct rigorous proof-of-concept testing that mirrors the production environment to ensure performance matches expectations. If performance is up to par, an appliance may be a smart choice, especially for organizations that need to reduce expenses for deployment and maintenance.

Nimaya Inc.

Q What can I accomplish with enterprise data mash-ups (EDM) that I can't from application integration?

A Enterprise data mash-ups (EDM) are lightweight, Web-based SOA applications that leave source data in its original location and state. They do all the work on the fly, usually without the need for "big iron." The smaller footprint of EDM enables companies to quickly construct new dynamic and flexible applications for specific tasks by assembling ingredients from existing applications, Web services, and databases. This empowers companies to solve problems incrementally, realizing ROI gains in small, measurable steps and avoiding the huge burden of a larger enterprise development project that could take months and millions. Management, developers, and end users all benefit from this approach.

ANALYST VIEWPOINT

Enterprise data mash-ups represent an intriguing convergence between bare-knuckle data integration and the rich interactivity of Web 2.0 Internet sites. Using such standards as JavaScript, DHTML, and XML for rapid development and integration in a service-oriented architecture, mash-ups provide a virtual, noninvasive tool to enrich certain applications, such as employee portals or supplier extranets, with dynamic structured or unstructured content without impacting source systems performance. While mash-ups are gaining popularity in certain niches, security vulnerabilities and relative immaturity means that they should not (at least for now) be considered as a replacement for a proven and robust data management infrastructure underpinning mission-critical systems.

Oracle

Q What is the value of prebuilt analytic applications for data integration and business intelligence?

A The use of analytic applications is a rapidly growing trend as organizations look to deploy BI more broadly and in a more integrated fashion with operational applications and processes. A key part of the value of analytic applications is the prebuilt ETL adapters that are typically included, enabling organizations to very quickly and easily deploy a data warehouse populated with data from common enterprise applications such as Oracle, Siebel, and SAP. Analytic applications also offer significant value when it comes time to upgrade, with utilities to help organizations upgrade the analytic environment in concert with the operational applications running in parallel.

ANALYST VIEWPOINT

Analytic applications can offer the best of both worlds—prebuilt functionality that zeroes in on issues common to a functional area (e.g., supply chain or finance) or a particular industry, as well as customization options that enable developers to tailor the application to an organization's unique needs. As a result, they have become a popular alternative to both vanilla query and reporting tools and to building an application from scratch. While customization is always required, analytic applications can accelerate time to value for both developers and end users, with prebuilt ETL connectivity and business-focused workflows and other analytic functionality geared for the business side.

Petris Technology, Inc.

Q What can owners of long-lived assets do to maintain knowledge across the generations of technology, science, and staffing?

A Managing the information associated with these extremely large and long-lived assets poses unprecedented challenges—but also opportunities. Robust data management framework and process are required to ensure that relevant information is captured and managed for future use. A workflow that automates many of these processes removes the need to manually handle inputs, while the collected insights of experts provide built-in training for the next generation who are just learning the ropes. The long life of a field will mean that technology, scientific concepts, and analytical approaches will evolve during its life and adjustments must be made. A solid data management approach and integrated workflow will ensure these advancements can be applied to distill more knowledge from old data.

ANALYST VIEWPOINT

The question has more to do with political will and strategic vision than with technology. Organizations in highly technical industries do face unique challenges with massive and complex data volumes, the need to preserve knowledge across generations, and the certainty that information will only continue to grow. On the other hand, data management systems and storage capacity have matured to help meet the challenge. For one, the discipline of information lifecycle management (ILM) provides a framework to align the business value of data to its most appropriate and cost-effective storage medium. Implementing an overarching information system like ILM can be a massive challenge, but it's also increasingly crucial for scientific and research organizations.

Pitney Bowes Group 1 Software

Q Why has interest in improving master data management surged in recent years?

A In many respects, the interest in master data management (MDM) is a result of lessons learned during struggles to implement effective CRM and ERP applications, which were intended to provide a 360-degree view of a given customer. Many organizations come up with a fuzzy or distorted view because the underlying data used to generate it was of poor quality. Now businesses recognize the importance of high-quality data as the basis for any business intelligence programs that they implement. MDM can help align and maintain master data assets.

ANALYST VIEWPOINT

Data quality and consistency has long been the crazy uncle that no one in the family wanted to acknowledge. Organizations have historically had a limited understanding of the costs and inefficiencies associated with data discrepancies across multiple systems, as well as the return on investment (ROI) that can be derived from MDM and a single set of consistent data. Now, the maturation of MDM platforms and data quality technology have given enterprises a means of automating what would 10 years ago have been a costly and painstaking manual process. Publicized MDM success and the perceived opportunity costs of failing to address MDM also contribute to MDM's ascent on the IT radar screen.

Siperian, Inc.

Q Certain IT departments are still finding it challenging to build the right business case for master data management initiatives. How do you recommend creating a compelling business case?

A This is an important question—MDM is no longer exclusively an IT-driven project. MDM can provide significant business benefits, and gaining business sponsorship up front is critical. We recommend doing a quick ROI audit to identify the specific set of business drivers by function—whether in marketing, sales, contracting, procurement, or compliance. The key is to identify the hard cost savings associated with a MDM project that will capture the attention of business managers. For example, one manufacturer identified over \$11 million in cost savings from their sales operations over a five-year period, effectively paying for the MDM project. However, it is equally important not to leave out the soft costs, which are essential to sell the larger vision.

ANALYST VIEWPOINT

Most IT professionals, much less business sponsors, have little insight into the magnitude of data quality and inconsistency problems across multiple business units. A sound first step in building a business case for MDM is to document the extent of the problem through a data life-cycle audit (and in many cases, the problem will be greater than expected). A thorough assessment of the issue supplies a foundation to 1) document the cost of data discrepancies to the business, and 2) quantify business benefits and ROI that can be realized through MDM, ideally over a 5- or 10-year period. It may be helpful to document the larger industry trend toward a comprehensive data infrastructure—and the opportunity costs of falling behind.

Sybase

Q What are the most important criteria for creating a usable data aggregation environment?

A There are five requirements that stand out. A data aggregation environment should:

1. Support multiple data integration technology requirements, sharing common administration, design, and metadata.
2. Meet rigorous standards for ad hoc querying and data warehouse load performance, well beyond what can be delivered by traditional relational database systems.
3. Be significantly easier to deploy and maintain than what is possible through traditional database systems.
4. Enable near-linear user and data scalability to support thousands of users and terabytes of data and, in some cases, also the flexibility to allow multiple grades of SLAs.
5. Provide a cost-effective solution that drives data aggregators' top-line growth while containing data warehouse infrastructure costs.

ANALYST VIEWPOINT

Throughout the process, it's important to keep the overarching objective in mind—building measurable and sustainable business value, which can become diluted amid the complexity and time pressures of a large-scale data aggregation project. Key criteria supporting that principle include 1) executive-sponsored collaboration between business and IT, 2) improving the quality and consistency of enterprise data, 3) ensuring performance and scalability, and 4) implementing a standards-based system that may be rapidly and affordably extended to support future initiatives or mergers/acquisitions. Establishing a “center of excellence” is a proven way to coordinate competing priorities and share best practices to help ensure that business value is realized.

Syncsort Incorporated

Q How can I improve the quality of my data throughout my data integration project?

A A thorough data quality program includes two phases. First, data is either captured in a standard, error-proof way, or it is cleansed in preparation for loading. Then, data can be enhanced for further analysis. For example, demographic and/or lifestyle information can be added to customer records before the actual load. Enhancing data usually entails combining multiple sources of data, and this data is often held in multiple databases on disparate platforms. Using a data manipulation tool makes coordinating all of these different sources much easier.

ANALYST VIEWPOINT

With growing recognition of the risk that poor data quality poses to the business, many organizations are taking the smart step of incorporating data profiling and cleansing into broader data integration initiatives, often multisource integration into a warehouse or migration from legacy systems into a modern, Web-based application. Data profiling is a key first step that assesses the content and structure of data—an information reconnaissance mission that is prerequisite to thorough cleansing and reconciliation across multiple sources. Ideally, data quality is viewed not as a collateral project, but as integral to a broader integration initiative with appropriate scoping, tools, and resources allocated to help ensure its success.

CASE STUDY

MedSolutions: Improving Healthcare through Information Access and Streamlined Processes

Commentary by Steve Wise
CIO, MedSolutions

Company Overview

MedSolutions is a leading radiology management company that provides customized, proprietary management of high-tech imaging to a national client base.

Business Challenge

MedSolutions has recently undergone significant growth, which led to a dramatic increase in call center volumes. In addition, the company continuously strives for faster, better ways to work with physicians who request imaging studies, and to streamline the authorization process for them. Furthermore, MedSolutions was looking for a solution to provide members with increased transparency of information, and to support health plan customers in delivering this and other consumer-driven healthcare information. As a result, the organization's plans included a complete overhaul of its prior authorization and notification request process for imaging studies online.

Solution

To improve service delivery to physicians, support client health plan needs, and proactively manage increased demand on call center staff, MedSolutions executives decided to develop a self-service portal that would enable customers and other users to obtain immediate adjudication of imaging requests, and determine approvals or recommended clinical reviews.

To create the portal, MedSolutions partnered with Collaborative Consulting. The team expanded MedSolutions' proprietary neural modeling (a competency built in to the existing call center application), enabling the portal to immediately adjudicate most

"cases." The expansion was also designed to automate approvals to physicians' offices without human interaction.

As the project began, Collaborative sent a team of user experience design professionals to MedSolutions. The team studied the organization's business processes—from customer through call center and interactions with managed care companies—as well as its customer requirements and business rules. After evaluating the environment, Collaborative and MedSolutions technology professionals Web-enabled the process that providers (i.e., physicians) undertake to receive authorization, all the while mirroring the existing call center process to foster consistency.

Next, Collaborative introduced experienced project managers to keep the project on time and within budget. Other Collaborative contributors included a group of technology professionals with specific portal architecture and systems integration expertise, and several data warehouse architects. These experts synchronized three disparate data sources—claims, insurance eligibility, and member-imaging authorizations—making patient imaging information readily available to providers while increasing the portal's functionality significantly.

Benefits

The robust application offers "My Portal" functionality to members, providers, imaging centers, and health plans that want to conduct ad hoc reporting and customize services based on personal desires. Sophisticated scripting and application business rules support provide compre-

hensive member redirection to lower-cost, high-quality imaging centers. The portal also allows MedSolutions' health plan customers to enhance their brand via their own customer-facing Web sites, featuring information from MedSolutions' back-end databases. Other services are being delivered as the rollout continues. These include eligibility verification, claim status, electronic claims submission, and educational programs.

The self-service portal provides a clean and modern visual environment and a strong design that reflects accuracy, simplicity, timeliness, professionalism, and efficiency. These attributes adhere to MedSolutions' corporate design standards, as well as to its objectives of higher quality and lower diagnostic imaging costs for customers, and reduced call center volumes. In addition, the portal features an intuitive user interface, and delivers immediate access to imaging authorization status, claims, and provider profile information.

Further, it synchronizes patient claims, eligibility, and member imaging authorization data. For example, by integrating fax and mail technology with online processes, providers receive immediate authorizations for treatment and state regulated letters. For more clinically complex cases, detail clinical information or attached progress notes allow nurses and doctors to process cases efficiently.

From a data perspective, the benefits are abundant. The portal can quickly and easily provide reports to specific internal and external users without deploying software or establishing other means of delivery. Internally, data can be leveraged in numerous ways. For example:

- Marketing and product development teams gain insight into the products and services various customers use most often, those they value most, and services that might potentially add value in the future.
- The medical management team can compare patient profiles among people with similar conditions to determine if physicians order unnecessary tests or treatments.
- The sales team can identify opportunities to cross-sell by comparing demographics, behavioral data, product and service use, and survey information.

Externally, partners can use the portal to leverage data and enhance care. For instance:

- Health plans have access to key radiology utilization and cost data for their members, provider reporting, and a platform for private labeling radiology approval processes. With private labels, they can personalize and offer role-based views of core business functions and information to a variety of audiences.
- Providers have access to clinical guidelines and can review ordering patterns and benchmarks.
- Members have access to their imaging information, allowing them to make informed decisions regarding their treatment, including choosing providers based on quality ratings and out-of-pocket expenses.



Figure 1. The robust application offers “My Portal” functionality to members, providers, imaging centers, and health plans that want to conduct ad hoc reporting and customize services based on personal desires.

The portal’s architecture is leveraged as follows:

- **Reusability.** A services-based approach to components within the solution’s data layer facilitates the reuse of components by leveraging existing legacy systems, resulting in a consolidated, enterprise data layer.
- **Reliability.** The asynchronous messaging layer provides a consistent user experience, including the ability to continue to submit authorization requests even during back-end system outages.
- **Loose coupling.** Loose coupling between the business logic layer and data services insulates the business and presentation logic from database schema changes. This approach facilitates the implementation of a major data quality initiative and a new claims processing system with minimal impact to the self-service portal solution.
- **Composite application design.** By designing the presentation layer as a granular set of interoperating portlets, business processes and application flow can be tailored

for different user experiences, or can be based on specific client requirements.

Thus far, the portal has received widespread approval among medical professionals. In addition, it is streamlining call center volume internally, while its user-centric design and immediate access to radiology management data improves customer relationships. Critical first steps of understanding user needs and critical business functions have also paid off. MedSolutions has realized its adoption goals and recouped costs as a direct result of these critical first steps. As time goes on, patients, imaging facilities, and healthcare benefits administrators will benefit from its functionality, and customer satisfaction will continue to increase. ●

Leverage a Disciplined, Standard Approach to Achieve “One-Source” Truth

By Mark Holmes

National Operational Consulting Practice Director

John Williams

Senior Vice President, National Practice Director,
Collaborative Consulting

Today’s best-run companies can identify and repair problems and replicate best practices throughout the organization very quickly. Most of the time, these organizations measure and manage their supply chains with exceptional skill and vigilance.

Excellence in supply chain management means achieving superior levels of visibility into systems and processes, while optimizing the effectiveness of key performance indicators (KPIs). Companies can improve their visibility into systems and process execution by ensuring that their KPIs are well defined and based on accurate, timely data. When organizations accomplish these levels of introspection, they can create effective plans and quickly implement changes that make a real difference.

Here’s the catch: KPIs rely on multiple, usually disparate sources of data. These feature a wide range of quality and accuracy, making it difficult to draw precise conclusions about their intended measurements. In fact, defining the proper KPI and gathering its associated data is usually incredibly complex and challenging. And the more functions and divisions an organization has, the more complicated the task.

Consider a particular data value. An organization’s enterprise resource planning system, its third-party logistics system, and its general ledger may all provide this value. However, each could come by that value differently, which invites inconsistency. Figuring out how to rein in all that information, determine what it means, and represent it in an accurate and timely fashion is indeed a

tall order. How can a company achieve “one-source” truth?

The best way is to adhere to a common, disciplined approach to KPI development. Proper KPIs are the result of a mathematical equation used and integrated throughout the enterprise, and adopted by trading partners. This level of standardization helps KPIs link with an effective data architecture that enables timely, accurate visibility. When that occurs, underperforming processes are highlighted very quickly, as are those that ought to be implemented throughout the enterprise.

In addition, to optimize the effectiveness of its KPIs, a company must identify those that are most critical, i.e., the ones that drive the business. Usually, this is a relatively small but highly important group.

Because most companies have only a few critical KPIs, it makes sense to begin KPI enhancement initiatives with them. Doing so, however, is complex; a company must uncover, organize, and expose the data that creates visibility into KPIs.

Other important elements of KPI visibility include:

- **Data source.** KPIs are most effective when all functions and divisions within an organization “pull” a particular value from a standardized source, such as an ERP system. When the same value is extracted from different sources, inconsistencies and inaccuracies creep in.
- **Data users.** If only a small group of executives needs to review KPI data,

a relatively simple application can be developed to share that information. If, however, several manufacturing plants are also heavy KPI data users, and a global executive team requires the information, a portal may be more appropriate.

- **KPI value.** KPIs must align with business objectives. In addition, senior executives must ensure their adoption throughout the enterprise.
- **Measurement time frames.** By determining how often to measure KPIs, a company can figure out which data sources it must access easily. Of course, measurement times vary greatly from company to company. An e-fulfillment organization will want its “out-of-stock” KPI measured daily; another enterprise may need that metric only weekly or monthly.

KPIs require a sound definition, disciplined data gathering, and continuous alignment of that data with business objectives. By adhering to a rigorous plan and maintaining data discipline, companies can create extremely robust KPIs, and access powerful, accurate, optimized assessments of their performance. ●

For free white papers on this topic, [download](#) “Data Governance: Developing a Foundation for Long-Term Stability and Strategic Business Advantage” or “Data Visibility and Supply Chain Intelligence: Creating an Optimal KPI Model,” or [click here](#) for more information about Collaborative Consulting.

LESSON FROM THE EXPERTS

Master Data Management: Extracting Value From Your Most Important Intangible Asset

By **Seth Halpern**

Senior Principal of Value Engineering, SAP

In a 2006 best practices survey of the Americas' SAP Users' Group, 93 percent of respondents experienced data management issues during their most recent projects. And data management was identified as the root cause of problems in process improvement projects. Part of the problem stems from the fact that many organizations believe they are already using master data. But the reality is that they are operating within the confines of disconnected silos of data—the information contained in multiple systems, applications, and spreadsheets. Once master data is generated and trapped in silos, inaccurate and inconsistent information is perpetuated throughout the organization—and beyond. This creates an incomplete view of the business and limits your ability to aggregate and distribute data, which hampers even the most carefully planned and executed initiatives. Achieving master data consistency for all systems within a distributed IT environment has traditionally proven difficult.

An organization's primary data entities generally include customers, products, employees, and suppliers. As the volume of data in any or all of these grows, the importance and complexity of managing the data increases. For example, wholesale distribution companies may sell hundreds of thousands or even millions of different products. Maintaining accurate data on each of these products, SKUs, and components across the inventory is challenging enough, but imagine the problems across suppliers, customers, and third parties when naming conventions are inconsistent.

Effective MDM offers a number of benefits to a wide range of companies—particularly those with large volumes of data used for

a variety of purposes by multiple organizations. They include operational efficiencies, enhanced revenue opportunities, better insight into business operations, and tighter compliance with regulatory requirements—helping avoid fines, internal misbehavior, and the potential for losses in shareholder value. In virtually every respect, increased attention to master data management can lead to quantifiable increases in business value.

Market and operating incentives, combined with the advent of innovative technology, make a compelling case for deploying a master data management program. This is especially apparent when you consider the potential for improvement across all business processes.

Overcoming Barriers to MDM Excellence

To achieve effective master data management and improve operating performance, you must adopt a solution that addresses the following three elements:

Master data consolidation: Matching, normalizing, cleansing, and storing master data imported from client systems. The principal activities of master data consolidation are:

- Identifying identical or similar objects spread across local systems
- Building consolidated master data
- Providing ID mapping for unified, company-wide analytics and reporting

Master data harmonization ensures that master data is synchronized across heterogeneous system landscapes. Extending the scope of master data consolidation, harmonization also encompasses the distribution of consolidated, globally relevant information, and the enrichment of client application systems with locally relevant information.

Central master data management speaks to the maintenance and storage of master data and the development of distribution mechanisms for delivering master data to the systems that need it. This activity differs from master data harmonization in that master data is created centrally using a rich client.

You can then interactively distribute information to clients as required.

The ideal solution integrates seamlessly with your organization's existing infrastructure and those of your partners. Additionally, the solution is intelligent enough to ensure ongoing harmony of accurate and up-to-date information from disparate sources and is readily accessible to ensure it supports the needs of the entire business ecosystem.

Leading-edge technology helps you streamline and improve the aggregation of master data from disparate sources. The ideal solution manages the entire process, including deduplication and normalization, ID mapping, matching and merging, staging, change tracking, interactive data-quality analysis, and ad hoc consolidation. You can then analyze consolidated data using a business intelligence solution. Ideally users experience near-real-time search performance, with multiple search mechanisms with every dimension interlinked. You should be able to search an entire repository easily with any item or group of items, and partial strings and equivalents should be indexed to increase positive results. Performance should be measured in milliseconds—even when managing repositories containing millions of records.

The Case for Master Data Management

Ultimately, master data management, when done correctly, enables reliable cross-system, enterprisewide business processes and analytics, ensuring that everyone involved in the process has access to the same information and knowledge. A solution that enables the consolidation of master data, as well as the availability and free flow of consistent data across system boundaries, offers the most promising opportunity for improving business processes and a decisive competitive advantage. ●

Learn more: <http://www.sap.com/usa/platform/netweaver/components/mdm/index.epx>

[Click here to request more information about SAP.](#)

CASE STUDY

LexisNexis: Enhancing Sales Campaign Support and Improving Data Quality

Steve Iddings

Manager of Application, Integration and Self-Services, LexisNexis

The Business

LexisNexis (www.lexisnexis.com) is a leader in comprehensive and authoritative legal, news, and business information services. In addition to its Web-based Lexis and Nexis research services, the company includes some of the world's most respected legal publishers.

The Challenge

LexisNexis wanted to implement a master data management (MDM) solution that would enable its North American Legal Market (NALM) and U.S. Corporate and Public Markets (USCPM) groups to be more effective in the way they serve customers based on an aggregated view of over 5 million subscribers. The company wanted to establish a front-office environment to uniquely identify customers and improve the customer experience and effectiveness of marketing and sales functions.

The front-office environment was expected to make customer data available to support all users of those systems—sales, marketing, and support—and synchronize customer information maintained in the front office with the back office. Specifically, the company needed to aggregate customer master data from seven systems, including two sales-force automation applications, marketing data warehouses, business intelligence warehouses, and external data sources to provide better data to the front-office system for enhanced sales campaign support, and account and contact management.

At the same time, the company was upgrading the Siebel customer relationship management (CRM) system to leverage new capabilities. However, this would not address the fundamental problem of poor data reli-

ability, since customer data was entered and managed in multiple systems across the organization. As it turned out, the previous implementation of Siebel had failed to deliver the full value of investments due to unconsolidated customer data dispersed across systems. To address this, LexisNexis required a comprehensive master data management solution that would provide a reliable aggregated view of more than 5 million subscribers.

The Siperian Solution

LexisNexis first considered the option of building a solution themselves. However, after careful evaluation of the requirements, the company realized the complexity of the needed solution and decided to invest in an MDM solution. After an exhaustive analysis of all available offerings, LexisNexis chose Siperian MDM Hub for its built-in data stewardship capabilities, its data model flexibility, the overall completeness of the product, and its ability to support any master data type and integrate with any data source or application. Siperian also has one of the most sophisticated, code-free match, merge, and deduplication capabilities, with proven rapid implementation.

“Without Siperian MDM Hub to improve overall data reliability, we could not have achieved sufficient end-user adoption to realize the value of the substantial investment we'd made in our SFA systems.”

Steve Iddings, LexisNexis

In conjunction with the Siperian team, LexisNexis was able to implement the initial phase of the solution within eight months. Master data was integrated from seven different sources to create a hub of reliable master customer data as well as other data entities including location, accounts, firms, agreements, and contacts. Using Siperian MDM Hub, the company is able to create unique and actionable views of their vast customer base, synchronize this reliable data with front- and back-office systems in near-real-



time, and leverage a common data model throughout. This enables the company to align territories for optional effectiveness, and consolidate accounts and contacts on a nightly basis. This is a significant improvement, as consolidation used to take one month—resulting in poor communication among business units and lost up-sell opportunities. LexisNexis can now also identify which products their customers own across business units to minimize unnecessary or duplicate customer outreach efforts and identify possible product bundles based on actual user preferences.

Following the success of the initial implementation, LexisNexis extended the use of Siperian MDM Hub to include its online bookstore. By integrating the master data management hub into this self-service portal, more than half of all book orders are now fully automated, and greater automation is achieved for the remaining orders with the superior matching capabilities. Planning is already underway to further leverage Siperian to identify and manage product data to create a unified view of customers, products, and locations.

The Results

Siperian MDM Hub has become the cornerstone architecture for the company's CRM initiatives. The solution enabled LexisNexis to provide a more accessible and reliable customer master data store to all of its business users in sales, marketing, and support. As a result, the company realized continuous data improvements and an increased emphasis and company-wide focus on data quality and data stewardship. More importantly, LexisNexis was able to enhance sales campaign support and increase revenue. With a reliable master data foundation for all of their applications, sales personnel are now able to avoid missed sales opportunities, and business users have access to the most trustworthy enterprise customer information. ●

LESSON FROM THE EXPERTS

Master Data Management: Myth of the Mega-Vendor Platform

By **Anurag Wadehra**

Vice President, Marketing and Product Management, Siperian, Inc.

Industry analyst firm IDC is predicting that the master data management (MDM) market will grow to \$10.4 billion by the year 2009, with a compound annual growth rate of 13.8 percent. Other analyst firms such as Gartner and Forrester claim that MDM provides significant business benefits and is a critical foundation for managing enterprise information assets. Not surprisingly, the market hype from the mega-vendor triad (IBM, SAP, and Oracle) has been deafening as each attempts to position its single enterprisewide platform as a complete solution to MDM strategy. Despite the flurry of activity, no one is talking about MDM in the same way, and there is no consensus on the crucial questions that CIOs and executive management teams want to address. While MDM has come to the fore as a critical area of data management practice with different data domains, MDM requirements have not yet coalesced into a coherent market.

Is a Single Vendor Platform Necessary?

At the organizational level, MDM is a cohesive strategy for managing all master data domains, whether product, customer, employee, asset, or financial. Each of these master data domains differs greatly. For instance, customer master data often originates from multiple sources, including several from outside the company. Customer master data is often structured and well understood, while typically following a “business party” model. On the other hand, product master data is usually generated internally and shared externally among suppliers. The characteristics of product master data—such as in a product catalog or an item master—are both structured and unstructured, while requiring a hierarchical data model.

It is being debated whether any mega-vendor today can handle these differing requirements in a single integrated MDM platform. But here is the more pertinent question to ask: is centralizing on a single MDM vendor even necessary? If the service-oriented architecture (SOA) promises of mega-vendors are indeed true, then why not select different best-of-breed solutions that work together to deliver the most suitable solution? Shouldn't each MDM solution leverage your past investments in data integration infrastructure, legacy data hubs, and external data sources? Perhaps the most critical question to ask is: where is your organization most likely to derive business value from a MDM platform?

The Path Forward: Adaptive Master Data Management

Many companies are finding that the simplest route to move forward with MDM is to initially deploy one master data domain for a specific business need, and then extend to other data domains over time. Which data domain to start with will differ by industry. For example, it may be customer data in high-tech, doctor, or hospital data for a pharmaceutical company, counterparty reference data in institutional banking, or product data in retail.

By starting with one data domain, companies are able to achieve immediate ROI in an identified area, typically in less than six months' time. However, the key to this approach is to select an adaptive MDM platform that can easily be extended to different data domains over time to meet future requirements. An adaptive MDM platform should not only support SOA and allow you to coexist with existing



data hubs, data sources, and the larger application platforms, but also allow you to evolve the architectural style over time across the organization to implement a comprehensive master data management strategy.

Remember, MDM is a strategic process, and while mega-vendors are vying for dominance, each master data domain presents unique challenges. An adaptive approach is necessary to meet these challenges and enable you to start down the path to MDM without compromising your ability to evolve toward an enterprisewide master data management platform. ●

For a free white paper on this topic, [download](#) "How to Sell Skeptics on the ROI of Master Data Management," or [click here](#) for more information about Siperian, Inc.

CASE STUDY

Gaining New Insights with Master Data Management

Commentary by Linda Scott
Vice President, First Data Corporation
Michelle Compton
IT Project Manager, Chesapeake Energy



Enterprise Data Management Works at First Data Corporation

First Data Corporation (FDC) is a worldwide provider of electronic commerce and payment solutions. As the result of merger and acquisition activity over many years, the company is composed of multiple businesses—each with its own data, data definitions, and business processes. In this fast-paced business, the intention was to go back after the fact and make data and processes consistent. However, what the company was a single master data management system that delivered “one version of the truth” and a single consistent set of data management processes—an enterprise master data management system.

The first phase in FDC’s enterprise data management initiative was to implement Hyperion Master Data Management to deliver a central repository for enterprise-wide reporting hierarchies and master data, including a company-wide common chart of accounts and all associated attributes and properties. In tandem, FDC established a data governance structure and processes to create enterprisewide master data consistency and control.

FDC now has a consistent language and the ability to see and compare financial and operational results between business units “apples-to-apples.” With “one version of the truth,” employees no longer have dis-

cussions about data accuracy, and instead focus on data analysis and reporting. The central finance team gained consistent processes, consistent consolidations, and enterprisewide data transparency, while reducing the cost of reworking and reconciling data, reports, and key performance indicators (KPIs).

The entire enterprise is working more collaboratively on common systems and goals. FDC is proof that crossfunctional teams comprising business unit knowledge workers and IT technical staff can work together and deliver a data management system that works. ●



Chesapeake Energy Strikes it Rich with Hyperion Master Data Management and GIS Mapping

Chesapeake Energy relies on maps and spatial information to fuel success. Its systems and processes must be flexible to keep pace with change.

Chesapeake Energy uses a critical management tool, The Team Table, to identify employees that are involved at every level in a land lease or well. The table identifies points of contact for each geopolitical or operational area. The Team Table was originally created and maintained in a spreadsheet, which required constant, manual updates resulting from employee assignment changes, hierarchy changes, and adding new wells.

Today, business users record, track, and update a Hyperion Master Data Manage-

ment application with personnel assignment changes, which links with a GIS mapping application to deliver spatial representation of teams and territories. This new approach takes information that historically was defined by a map reference—but not represented spatially—and delivers it in the form of a map.

This scalable, easy-to-use system provides accurate and up-to-date information for Chesapeake Energy’s 5,000 employees. Now, anyone in the enterprise can identify an area on a map and immediately retrieve the team associated with that area.

The company has seen tremendous value, both in immediate returns and in future opportunities:

- Faster access to data for decisions and analytical options
- Freedom to build teams according to where oil and gas reserves are, instead of by traditional political map boundaries
- Faster and more accurate hierarchy data
- Enormous time savings for data maintenance and reduced effort in role assignments
- Ability to look back historically to help assess changes in production
- Increased security and audit trails not typically available through a spreadsheet

Many benefits result from Chesapeake Energy being able to perform tasks they could not do before. During the past 11 years, staff size increased by 500 percent and the company has 40 times more wells. The previous manual system couldn’t keep pace with that tremendous growth. This innovative solution merged needed factual data with maps to lead the way for even more growth in the future. ●

LESSON FROM THE EXPERTS

Master Data Management and Business Performance Management

By **José Villacís**

Senior Product Marketing Manager, Hyperion Solutions Corporation

A recent TDWI survey revealed that 51 percent of respondent organizations use a master data management (MDM) solution to support business intelligence (BI) and transactional applications, and 32 percent use it for BI alone.

Moreover, 83 percent of respondents reported that their organization had suffered problems due to poor master data; specifically, 81 percent cited inaccurate reporting as a problem.

But bad reports aren't the only potential issue if you don't have your master data under control.

When it comes to business performance management (BPM), you're not working just with operational reports, but also with financial information, analytics, KPIs, etc. This information is usually contained in several data stores, such as a data warehouse or multiple data marts, and in applications for financial consolidation, planning and budgeting, strategic scorecards, analytic applications, BI tools, and, of course, transactional and other systems that feed the BPM layer.

Under this scenario, the master data problem evolves from simply trying to centralize and synchronize master data from a system of record, into a problem of how to effectively manage changes in master data. This is because the frequency of change, the nature of the change, the type of downstream system, the number of attributes, and the applicable business rules can vary widely. Most important, these changes involve different departments and stakeholders, depending on the type of master data (employee, vendor, customer, product, financial, etc.).

Transactional or operational MDM tools typically ensure that the right product, supply chain, or customer master data exists in ERP, SCM, CRM, or other transactional system modules. This information needs to be readily available, usually in real time, to ensure the successful operation of these systems. The key requirement is consistency and timeliness.

But transactional MDM systems are not built to support the master data management needs of BI and BPM—the challenge is managing the changes that are required in master data structures over time. There are companies today that process thousands of changes per month in their master data structures!

MDM solutions that can effectively support BPM (sometimes called Analytic MDM), while complementary to transactional MDM solutions, must rely heavily on highly flexible but robust business rules engines that can handle the varied nature of changes in BPM-related master data. A key requirement is business user empowerment; otherwise, IT remains a bottleneck. Today most IT organizations still use spreadsheets, messages, and calls to confirm and agree on changes with business users before applying them. So the overhead cost and risk of errors from this manual, iterative process are high.

Business users know why and how changes in master data and reporting structures should be applied. Whether set off by typical business decisions, or by disruptive events like mergers, acquisitions, or reorganizations, it is business organizations, not IT, who initiate change. For years we've predicated in the BI industry that IT should empower business users to do their own analyses and reporting, allowing IT to focus on more productive

tasks such as user-enablement and systems administration and controls. MDM responsibility for performance management is analogous; business users should be responsible for making changes and maintaining master data, while IT enforces business rules and manages the environment.

So an MDM application must be built for end users—not only IT—with an intuitive user interface that includes all the functionality required for managing master data changes and the organization's master data lifecycle. For example, it must have real-time validations, immediate user feedback, and approval levels to control changes. In addition, it must provide robust auditing, versioning, rollback, and integrity check capabilities for managing the environment. Robust import and blending, as well as export features, are necessary, as the type, frequency, and format of updates will vary widely.

Managing master data is essential for ensuring the quality and integrity of the information in our organizations. Ensuring that transactional systems share the correct master data in a timely manner is critical. But making sure that the BPM layer or “management system” layer provides the right information for decision makers is a priority, too.

What's important is to identify the biggest pains first, then determine the right MDM solution(s). Start small, but outline a path to evolve your MDM solution to the enterprise level. ●

For a free white paper on this topic, [download](#) “The Business Case for Information Management,” or [click here](#) for more information about Hyperion Solutions Corporation.

BI Search and Text Analytics

NEW ADDITIONS TO THE BI TECHNOLOGY STACK

BY PHILIP RUSSOM

QUANTIFYING THE DATA CONTINUUM

Before drilling into BI search and text analytics, we need to review the spectrum of available data sources. After all, the “data continuum” has direct import on the scope of reports and other documents indexed by search or mined by text analytics.

The data continuum breaks into three broad areas.

- **Structured data.** At one extreme of the data continuum, structured data is commonly found in database management systems (DBMSs) of various types.
- **Unstructured data.** The other extreme includes documents of mostly natural-language text, like word-processing files, e-mail, and text fields from databases or applications.
- **Semi-structured data.** The area between the two extremes includes semi-structured data in spreadsheets, flat files in record format, RSS feeds, and XML documents. Many of these media are used with cross-enterprise data-exchange standards like ACORD, EDI, HL7, NACHA, and SWIFT.

Some data sources are hybrids that are hard to categorize. Despite the three broad types of data sources, the continuum includes sources that can manage both structured and unstructured data. For example, a row in a database table has a well-defined record structure that defines fields of mostly numeric data types. Yet, the same record may also have fields that are character data types, like text fields or binary large objects (BLOBs). Likewise, a report may contain structured data (or a query that fetches structured data), as well as report metadata and text in headings that can be searched. RSS feeds are especially problematic, since they can transport a

variety of information, ranging from prose (unstructured) to transactions (semi-structured).

In recent years, market research conducted by various software vendors and consulting firms has attempted to quantify the relative percentage split between structured and unstructured data in the average user organization. Most estimates name unstructured data the unqualified winner at 80–85%, leaving structured data in a distant second place at 15–20%.

However, TDWI Research finds that unstructured data is not as overwhelming in volume as previously thought. In an Internet survey conducted in late 2006, TDWI asked each respondent to estimate “the approximate percentages for structured, semi-structured, and unstructured data across your entire organization.” (See the top bar in Figure 1.) Averaging the responses to the survey puts structured data in first place at 47%, trailed by unstructured (31%) and semi-structured data (22%). Even if we fold semi-structured data into the unstructured data category, the sum (53%) falls far short of the 80–85% mark claimed by other research organizations. The discrepancy is probably due to the fact that TDWI surveyed data management professionals who deal mostly with structured data and rarely with unstructured data. All survey populations have a bias, as this one does from daily exposure to structured data. Yet, the message from TDWI’s survey is that unstructured data is not as voluminous as some claim.

Now that we have a new and different quantification of the unstructured segment of the data continuum, what should we do about it? We should all pare down our claims about unstructured data volumes, but we should *not* change our conclusions about what needs to be done. In other words,

UNSTRUCTURED AND STRUCTURED DATA IN WAREHOUSES TODAY

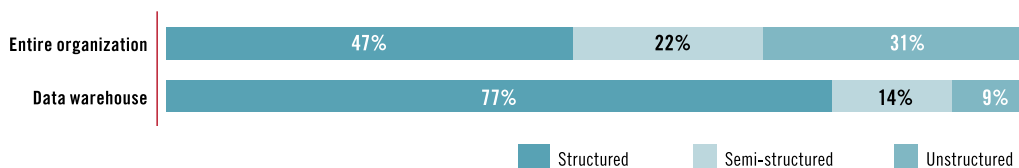


Figure 1. Little unstructured or semi-structured data makes its way into data warehouses today. Based on 370 respondents.

regardless of how the numbers add up, we all know that the average user organization has a mass of textual information that BI and DW technologies and business processes are ignoring. And this needs to change.

Why can't data warehousing professionals go on ignoring unstructured data? Among the many good reasons, two stand out:

- The view of corporate performance seen from a data warehouse is incomplete unless it represents (in a structured way) facts discovered in unstructured and semi-structured data.
- BI platforms today commonly manage thousands of reports, and techniques borrowed from unstructured data management (i.e., search) can make reports a lot more accessible.

To quantify the situation, TDWI asked each survey respondent to estimate “the approximate percentages for structured, semi-structured, and unstructured data feeding into your organization’s data warehouse or BI processes.” (See the bottom bar in Figure 1.) The survey responses reveal that structured data accounts for a whopping 77% of data in the average data warehouse or other BI data store, darkly overshadowing semi-structured (14%) and unstructured data (9%). Indeed, little data originating in unstructured or semi-structured form makes its way into data warehouses today, despite large quantities of it elsewhere in an organization. (Figure 1 compares these.)

The dearth of unstructured data in the warehouse isn't surprising, considering that almost all best practices in data warehouse modeling demand structured data. Likewise, we analyze and report off of data warehouse data using tools that see data only through the eyes of SQL, which in turn demands data in relational or multidimensional structures. As we'll see in detail later in this report, you have to impose structure on unstructured data before it's usable with a BI/DW technology stack.

NEW DATA WAREHOUSE SOURCES FROM THE DATA CONTINUUM

As we've seen, the data continuum divides into three broad segments for structured, semi-structured, and unstructured data. In turn, each of these segments is populated by various types of systems, files, and documents that can serve as data sources for a data warehouse or other BI solution. These range from flat files, to databases, to XML documents, to e-mail, and so on.

To understand which of these are feeding data into data warehouses today—and in the near future—TDWI asked, “Which types of data and source systems feed your data warehouse?” Survey respondents selected those in use today, as well as those they anticipate using in three years. Figure 2 charts survey responses for both today and the future; it also calculates the expected rate of change (or “delta”). Judging by users' responses to this question, the kinds of data sources for the average data warehouse will change dramatically in the next few years:

- **Unstructured data sources will soon be more common for data warehouse feeds.** The survey predicts the greatest increases with technologies that convey natural language information in text (aka unstructured data), like voice recognition (up 81% in three years), wikis (81%), content management systems (72%), taxonomies (70%), instant messaging (69%), and RSS feeds (68%). Admittedly, some of these show a high rate of change because they're starting from almost nothing, as with voice recognition and wikis (11% and 12% today).
- **Semi-structured data sources will increase moderately.** This includes stalwarts like XML and EDI documents (up 32% and 18% in three years, respectively). The new kid on the block is the RSS feed, which contains both semi- and unstructured data. Most RSS feeds transport prose (unstructured data as text), but are beginning to carry transactions as semi-structured data in markup documents. Either way, 22% of survey respondents claim that their data warehouse accepts RSS feeds today, and 90% anticipate integrating data from RSS feeds in three years. This makes sense, because RSS feeds operate in near real time over the Web, and many organizations

**WHICH TYPES OF DATA AND SOURCE SYSTEMS FEED YOUR DATA WAREHOUSE?
(SELECT ALL THAT APPLY FOR BOTH TODAY AND IN THREE YEARS.)**

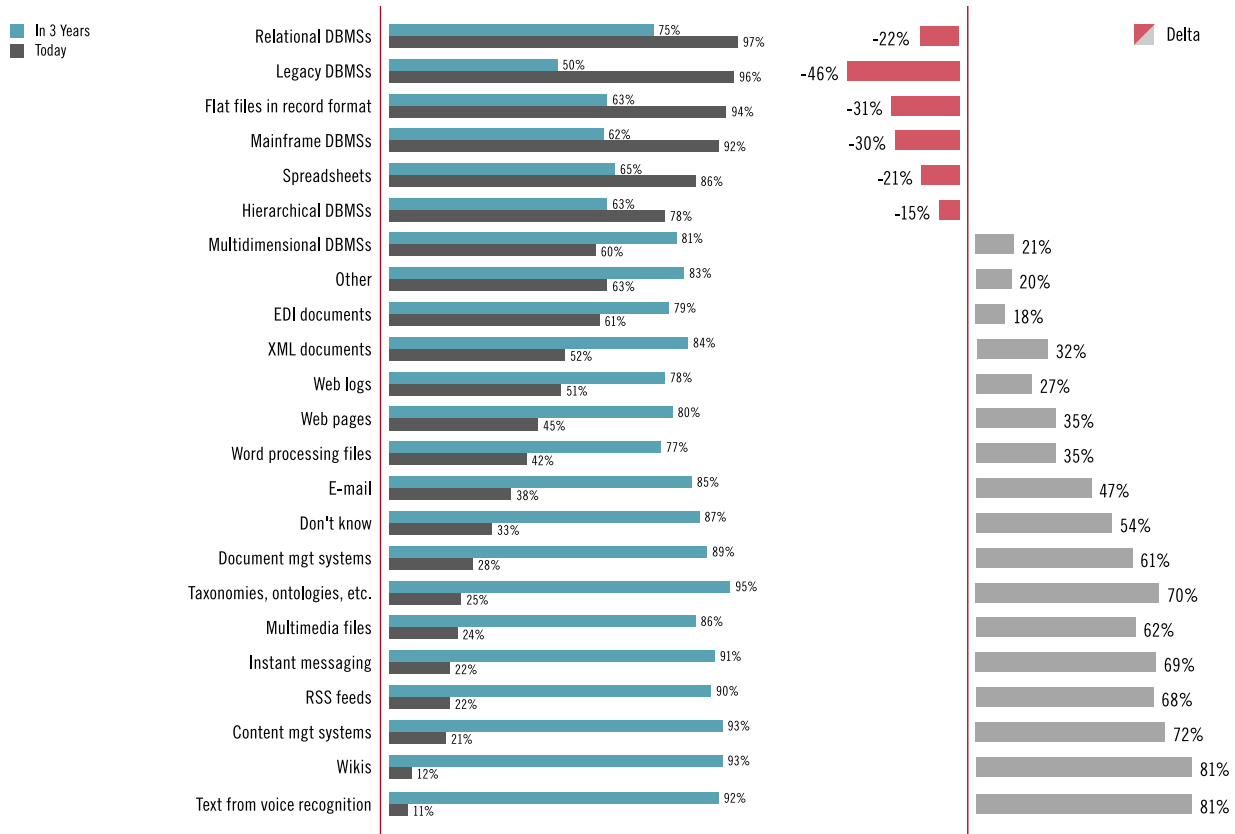


Figure 2. Based on 370 respondents.

are looking for faster and broader ways to deliver alerts, time-sensitive data, and transactions.

- Miscellaneous unstructured sources will increase moderately, too.** These are mostly files containing text, like e-mail (up 47% in three years), word-processing files (35%), Web pages (35%), and Web logs (27%). Their increase will be moderate because they're already established.
- Some sources of structured data may decline, but the category will keep its hegemony.** Survey respondents anticipate reducing data extraction from various older types of database management systems (DBMSs), namely those that are hierarchical (-15% in three years), mainframe (-30%), legacy (-46%), and flat files in record format (-31%). Indeed, these are legacy platforms that are ripe for retirement or migration. But survey respondents also anticipate extracting less data from spreadsheets (-21% in three years) and relational DBMSs (-22%). While the decline of legacy databases

as data warehouse sources seems plausible, TDWI Research is deeply skeptical about the decline in relational databases and spreadsheets claimed by survey respondents. Since these are so deeply ingrained in BI and in IT in general—and are spawning new instances constantly—their decline seems very unlikely.

The general trend—toward more unstructured data sources. Survey responses show that priorities along the data continuum will soon shift relative to data warehouse sources, with some data sources declining and others rising. Although respondents may have been overly optimistic about the rate of change they will embrace, the survey clearly signals a shift toward using more semi-structured and—especially—unstructured data sources. The trend is plotted conceptually in Figure 3, and the shift can be visualized as an increase in the types of data sources plotted in the middle or on the right side of the graph. Another way to see it is that the wide majority of data warehouse feeds today come from the left end of the graph. These won't go away, but instead will be joined incrementally by more data sources toward the right end.

DATA AND SOURCE TYPES PLOTTED ON THE DATA CONTINUUM

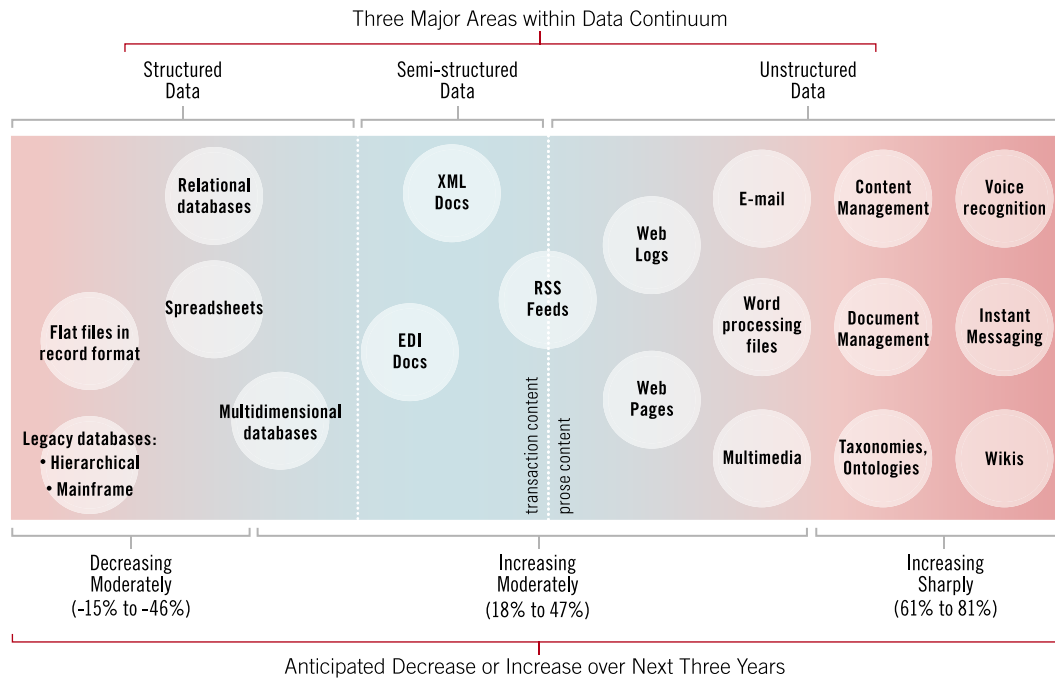


Figure 3. The data clearly signals a shift toward using more semi-structured and—especially—unstructured data sources.

RAMIFICATIONS OF INCREASING UNSTRUCTURED DATA SOURCES

The evolving list of data sources means changes for DW/BI practices. Data warehousing professionals should be aware of these and prepare for them:

- **Unstructured and semi-structured data must be transformed into structured data.** Note that sources of unstructured and semi-structured data will be increasingly tapped for data warehousing, but that doesn't mean that much of this raw data will actually go into a data warehouse. In most cases, this source data will need to be parsed for entity extraction or otherwise transformed into structures that are meaningful in a data warehouse or to a reporting tool.
- **Data integration will need to change substantially.** The wide majority of data integration routines for data warehousing today interface with structured data sources and transform the data accordingly before loading it into the data warehouse. Assuming that unstructured sources will increase, data integration for the data warehouse will need to reinvent itself in the next few years.
- **Data modeling could face a similar transformation, but not as extreme.** A few data models in data warehouses will require adjustments to accommodate

the structured data coming from unstructured data sources. Since the data is usually structured by the time it arrives in the data warehouse environment, adjustments should be slight. Similar adjustments are required when users want to copy unstructured data into a warehouse.

- **Training—and learning—are in order.** Data warehousing professionals currently have little or no experience with unstructured or semi-structured data sources. Likewise, experience is rare with search and text analytic tools. So additional training is needed, and—due to minimal experience—the learning curve will be long and flat. ●

Philip Russom is the senior manager of research and services at TDWI, where he oversees many of TDWI's research-oriented publications, services, and events. He can be reached at prussom@tdwi.org.

This article was excerpted from the full, 32-page report by the same name. You can download this and other TDWI Research free of charge at www.tdwi.org/research.

The report was sponsored by Business Objects, Cognos, Endeca, FAST, Hyperion Solutions Corporation, and Sybase, Inc.

CASE STUDY

Emerson Process Management Unifies Account Data for a Stronger Revenue Stream

Commentary by Nancy Rybeck

Customer Data Warehouse Strategy Architect,
Emerson Process Management

A Need for Coordination and Consolidation

Emerson Process Management, a leading global supplier of process-oriented products and services, is a \$3 billion operating unit of St. Louis-based Emerson, the *Fortune* 500 global technology and engineering firm.

Responding to global competition and a drive to increase profitability, Emerson Process Management has evolved from a product-centric organization to a customer-focused enterprise. Consequently, the organization required technology that could enhance its knowledge about the customers and partners of its global sales and manufacturing divisions. Understanding each customer fully and completely required accurate information instantly available to all reaches of the Emerson organization.

This was a daunting challenge, as Emerson Process Management's customer name and address data—representing over 75 business units—resided in multiple operational systems. It was found that 80 to 90 percent of those records could potentially be duplicates. To make matters worse, each division around the globe had its own method of managing its customer data. Each maintained its own database with a unique representation of customer names and addresses. This uncoordinated approach to customer relationship management could not support the firm's goals.

Gaining a Single Customer View

Emerson Process Management now uses Group 1 Software's flagship data quality solution, the Customer Data Quality (CDQ) Platform, which provides enterprisewide

correction, validation, and enhancement of customer, prospect, and supplier data. The CDQ Platform is helping Emerson Process Management meet its goal of obtaining a single view of each customer and partner. In a short time, 70 percent of its customer and partner address records had been identified as duplicates and eliminated. Also, nearly 85 percent of the unique business sites within its database had also been deduplicated.

Emerson is an excellent example of how improved data quality can help increase the agility of a company, thus contributing to a strong revenue stream, better operational efficiency, and increased shareholder value.

"As a long-time Group 1 customer, our decision to adopt the CDQ Platform was guided by the solution's unparalleled global address data support," said Nancy Rybeck, customer data warehouse strategy architect at Emerson Process Management. "And with a majority of our customers and partners doing business with many of our divisions, we required a solution such as the CDQ Platform that could perform powerful, user-defined record consolidation."

The CDQ Platform is an integral component of the organization's marketing efforts, most notably cross-selling and up-selling initiatives developed to expand penetration of

existing accounts. With higher-quality data and an increased customer focus, Emerson Process Management has seen tangible benefits, most significantly a double-digit percent increase in purchases per customer, a reduction in selling costs, and improved customer satisfaction.

Emerson is an excellent example of how improved data quality can help increase the agility of a company, thus contributing to a strong revenue stream, better operational efficiency, and increased shareholder value. They are one of thousands of progressive organizations that understand you are only as good as your data.

An Essential Component of Customer Data Quality

This solution combines leading-edge data quality functions in a single dataflow environment so you can make more informed business decisions. Sophisticated parsing, standardization, matching algorithms, and advanced address validation tools work together seamlessly so you can unify account data, eliminate redundancies, link members of the same household, and expand your customer knowledge. ●

LESSON FROM THE EXPERTS

Avoid the Pitfalls of Poor Data Quality

By **Bernie Gracy**

Vice President, Global Strategy,
Product Management and Marketing,
Pitney Bowes Group 1 Software

Without clean data, there is no CRM. Poor data quality can lead to serious business problems.

Several years ago, the introduction of new customer relationship management (CRM) systems from the world's largest software houses was heralded as the panacea for the wayward enterprise. The formula for the success of these systems seemed simple: interact with customers, collect valuable data, compile 360-degree views of individual customers, and utilize the information to build rich relationships that improve over time.

After a while, however, many enterprises found they were not getting the value from these systems that they expected, and they have arrived at similar conclusions: CRM systems are really just front-end graphical user interfaces (GUIs) and canned business processes (best practices) that rely on data. A superior CRM system does not guarantee data quality and is unable to generate return on investment on its own. It is the quality of the data that is fed into the system that makes all the difference.

Defining Data Quality

Data quality applies to more than just customer name and address data. It applies to product numbers and associated descriptions, part numbers and units of measure, medical procedure codes and patient identification numbers, telephone numbers, e-mail addresses, commodity codes, vendor numbers, and vehicle identification numbers, to name just a few.

Ensuring data quality requires:

- Understanding the nature of the data and the degree of "trusted authority" from which it is derived

- Understanding the intended use of that data
- Identifying factors that both determine and impact the data's fitness for use
- Establishing the policies, people, processes, and technologies to manage the quality of data

Given the current emphasis on the need to maintain a 360-degree view of the customer, data quality involves being able to link all of a given customer's records together—a task that can only be accomplished with identifiers for the records associated with each customer.

An Absolute Necessity

Poor data quality can impact an organization's ability to increase customer retention and loyalty, limit exposure, and increase operational efficiencies. For example, the inability to eliminate redundant name and address records can result in additional mail-order campaign costs, customer dissatisfaction—even legal concerns. Imprecise data on the total business conducted with a single vendor can result in missed opportunities for better rates with suppliers.

Recent regulatory and Homeland Security initiatives such as the U.S. Department of Treasury's Office of Foreign Assets Control (OFAC), Sarbanes-Oxley, the U.S.A. Patriot Act, and the Health Insurance Portability and Accountability Act (HIPAA) can quickly spur a company to establish a solid data foundation.

These regulations will cause even lagging organizations to recognize that an effective data quality program is quickly becoming a near-absolute requirement.

Steps for Improving Data Quality

Improving data quality and using data more strategically can be achieved by:

1. Conducting a data quality assessment to help recognize the severity of database quality issues.
2. Adopting a well-defined data governance plan, including a definition of who owns the data, who is authorized to access the data, and which specific standards should apply to the data.
3. Developing a corporate-wide agreement on data standards for master reference data that describes common business entities like products, customers, and suppliers.
4. Choosing a technology to serve as the backbone for preparation of relevant customer data that includes name and address and non-name and address cleansing, change-of-address processing, tax jurisdiction assignments, personalized messaging, tables and dictionaries, batch and real-time processing, and more. ●

For a free white paper on this topic, [download](#) "Data Quality: The Foundation of Operational Effectiveness," or [click here](#) for more information about Pitney Bowes Group 1 Software.

CASE STUDY

Intellidyn Uses Data Integration to Manage Industry-Leading Consumer Information

Commentary by Daniel Teachey

Corporate Communications Director, DataFlux

The Business

Intellidyn is an award-winning provider of the nation's most complete direct-response marketing and multi-channel database marketing and analytics solutions, delivering strategic insight and end-to-end marketing services to consumer businesses. In addition to compiling one of the industry's most "battle-proven" prospect databases, Intellidyn also maintains the largest, most in-depth repository of transaction, credit, demographic, and behavioral databases in the nation. The company provides unparalleled expertise in data analytics and integrated marketing to consistently boost ROI.

The Challenge

Building a successful consumer-centric data warehouse has been a challenge for many companies. The majority of these attempts have failed to meet expectations. Since data management was the very core of Intellidyn's business, failure was not an option for the company. Intellidyn's experience was that data warehousing and customer relationship management were not the sole domain of IT. Success was in allowing savvy database marketers to drive the IT solution.

More importantly, the competitive environment was becoming more fluid and dynamic. As consumers embraced new channels, such as online shopping and voice-activated transactions, these radically increased not only the volume of data that Intellidyn clients were handling, but also the speed at which it had to be processed. Intellidyn needed a data integration solution that could accurately and consistently identify individuals—and all data elements relating to those individuals—across 60 terabytes of data on more than 190 million U.S. customers in real time, while being transparent to "power users."

Since Intellidyn continuously updates its data warehouse from multiple sources as fresher information becomes available, the solution view had to span all disparate data sources to compile complete, accurate records of individuals, households, and addresses from over 2,000 unique attributes. This new platform also needed to integrate seamlessly with the company's existing IT platform. Additionally, to address the changing marketplace, Intellidyn would require an agile system that could answer complex queries in seconds, while converting, scoring, and arraying massive master file refreshes.

Intellidyn needed a data integration solution that could accurately and consistently identify individuals across 60 terabytes of data on more than 190 million U.S. customers in real time, while being transparent to "power users."

The Solution

Intellidyn implemented an enterprise data warehouse using DataFlux data management technology to handle data profiling, data quality, data integration, and data enrichment. DataFlux provided Intellidyn with the flexibility to create match codes based on fuzzy logic algorithms. The integration of this logic allowed Intellidyn to create different degrees of matching accuracy, doubling the company's matching efficiency.

DataFlux technology gave Intellidyn the ability to perform phonetic matching, cleanse common data errors, and standardize the

entire production process so that all consumer data could be linked consistently over time at three levels: individual, household, and address. This solution enables Intellidyn to generate more than 1 billion linking keys in less than 24 hours.

The Results

The DataFlux solution has contributed to an impressive increase in the efficiency of Intellidyn's data management. Access time to multiple master files decreased from 22 hours to 6. Intellidyn clients now have current, accurate, and reliable views of their customers within hours, instead of the days or even weeks that the same process once took. The enterprise data management solution offers Intellidyn a "360-degree view" of every consumer in the United States.

"It all comes down to what consumers demand from our marketing clients—'know me... or my loyalty drops to zero,'" said Peter Harvey, president of Intellidyn. "DataFlux provides the ability to build the customer-centric views enabling smart marketers to demonstrate to their customers that they do in fact 'know them.' In return, these marketers realize continuous boosts in performance. Without this critical data foundation, all other data warehousing and CRM functionality is compromised." ●

LESSON FROM THE EXPERTS

Data Integration Drives Healthy Customer Relationships

By **Tony Fisher**

President and CEO, DataFlux

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

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

Pervasive Problem

The IT staff pulled reports from its customer relationship management (CRM), enterprise resource planning (ERP), and billing and supply chain management systems. What they found reveals a distressing but pervasive problem that exists in most organizations.

Each list contained different, overlapping, and confusing views of the customer base. Some names were listed or spelled slightly differently; some addresses were a bit off. The end result was that this company could not create an accurate and inclusive list of customers affected by the loss of the distribution center.

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

Uncertainty about customers’ identities can severely compromise efforts to build stronger relationships. In today’s competitive marketplace, if customers don’t feel valued, they will take their business elsewhere. Customers are hard to acquire, but easy to lose.

Adding to the Customer Equation

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

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

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

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

The second component is identity logic—a crucial phase of any successful CDI effort. This determines whether customers listed in

different sources are indeed the same customer, and intelligently integrates customer information from multiple applications and databases.

The various records for Michael William Smith, Mike Smith, and Michael W. Smith, for example, are determined by identity logic to indeed be the same individual, provided other data points are similar. Companies can flag information for linking customers across applications and sources, and isolate the best data from multiple sources.

Building Lasting Customer Relationships

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

Thanks to CDI, organizations are developing healthier, more lasting relationships with their customers, and can market more intelligently—and profitably—to them. ●

For a free white paper on this topic, [download](#) “Semantics, Metadata and Identifying Master Data,” or [click here](#) for more information about DataFlux.

CASE STUDY

M&S Money Meets the Compliance Deadline for Basel II Through an Enterprisewide Data Quality Initiative

Commentary by Neil Hershaw,
Information Management Officer, M&S Money

The Challenge

M&S Money is a leading U.K. financial services company, a wholly owned subsidiary of global banking leader HSBC. It retains a close working partnership with UK-based Marks & Spencer and collaborates on customer activities. Like all financial services providers operating in the E.U., M&S Money was faced with the Basel II Capital Accord, requiring the tracking and reporting of ongoing exposure to credit, market, and operational risks. The compliance deadline was January 1, 2007. Basel II is a demanding piece of legislation that requires extensive expertise and sophisticated data management capabilities if the quality and integrity of data is to be consistently assured.

M&S Money therefore needed to deliver a data quality initiative that ensured data was managed according to a set specification throughout the whole information environment, imposing quality controls at many points throughout the architecture while retaining centralized control and management. Neil Hershaw, information management officer for M&S Money, saw this mandate as an opportunity to improve data quality enterprisewide. “Basel II’s compliance deadline presented us with both a challenge and an opportunity—delivering high-quality data to ensure successful risk management, but also improving our data to drive improvements in many other areas of our operations.”

In addition to the compliance requirements, M&S Money sought broader data management benefits. These centered on the

need to avoid potential customer frustration caused by inaccurate data, enabling the company to undertake business analysis and, therefore, sales activities with greater confidence. In selling to Marks & Spencer customers, M&S Money has adopted the policy that when it is not completely confident about data accuracy, and therefore a customer’s likely interest, it will not approach the customer at all, rather than risk any frustration. “We cannot ever risk upsetting a customer because of problems with our data,” Hershaw crystallized the issue.

“There was obviously the need to satisfy the data quality elements of Basel II, but also taking this enhanced overall approach to data management would give us greater confidence in our activities to create sales and therefore help drive competitive advantage in the market.”

Neil Hershaw, M&S Money

As part of the project to ensure Basel II compliance, therefore, M&S Money planned and delivered a data quality project to accomplish a number of operational benefits, namely the assurance of customer data accuracy, reduction of time taken to complete business analysis, and increased assuredness in the

quality of data from which business intelligence is created.

The Solution

M&S Money chose the Informatica PowerCenter data integration solution as the cornerstone of its project. Informatica Data Quality was specifically selected to assure the quality of data input and extraction from its Oracle data warehouse. Given the impending Basel II deadline, M&S Money created a specification in mid-2004 for a data quality initiative that would meet the regulatory requirements and the Financial Services Authority (FSA) mandates, as well as deliver a platform for enhanced organization-wide data quality management. Informatica was chosen because of PowerCenter’s technical attributes, its scope for customization, and its ability to integrate a multitude of disparate data feeds. Moreover, the Informatica team demonstrated a wealth of knowledge of how best to achieve the level of data quality required for Basel II compliance and brought the experience of managing similar scales of assignments for major banks and financial institutions. “Basel II sets the bar high on required data management practices. Compliance dictates the ability to correlate a significant history of consistent, accurate, and granular data. Equally, as an organization, we wanted to ensure we could always undertake quantitative measurement of our data to be utterly assured of its accuracy,” said Hershaw.

Constant Monitoring of Data Quality

The primary operational driver was that M&S Money needed to be constantly aware that its

MARKS & SPENCER

data was of the appropriate quality and that it could always demonstrate clear management of the data. Informatica Data Quality was deployed as a data quality management platform alongside the data warehouse to ensure that data feeds comply with the required quality levels, ensuring that BI data remains wholly accurate. Equally, the solution had to support M&S Money's new data management procedures, which were introduced as part of the project, so that any changes that could negatively impact quality would be immediately addressed and nullified.

Ensuring Data Quality from Third Parties

In creating this caliber of data quality management platform, the company also had to consider the need to assure the sustained accuracy of data derived from third-party feeds, such as those from insurance policy providers. Again, Informatica Data Quality provided the required support for these disparate data feeds, and therefore enabled all data quality to be managed within a single framework.

Regular Evaluation

The methodology for the project began with defining data quality rules for the relevant files and tables, coding those rules, running the data, then analyzing the data and creating an action plan for quality assurance. Once the project is completed, data quality levels are evaluated formally on a quarterly basis.

The Results

Basel II Compliance

The main driver for the business was com-

pliance with Basel II's stipulations; this was completed successfully a year ahead of the deadline. The company has created an assured level of data quality that has enabled it to undertake business analysis and financial reporting with greater confidence.

Forty Percent Reduction of the Analysis Cycle

A further gain has been the reduction of the analysis cycle by up to 40 percent, meaning faster comprehension and validation of information thanks to less time spent assessing the accuracy of the data.

Flexibility and Support

"We evaluated several potential solutions," Hershaw summarized, "and chose Informatica because they have vast experience in Basel II compliance and Informatica PowerCenter was the right solution for our needs. It offered ease of use, support for our existing IT environment, and the ability to be customized to meet our precise requirements. We needed to ensure that all data feeds, be they internal or external in origin, were in line with our quality specifications. The Informatica platform has given us the ability to provide that assurance to our business and to enable our analysts to create rules for how the data should be treated, without having to consult a program or make changes to our systems."

An Engine to Drive Business Development

M&S Money's IT team is now able to provide quantitative measurement of the data held by the organization, which was a broader goal of the initiative. For a financial services company, that resource has proven to be invaluable in providing the "engine room"

that drives business development. After Informatica Data Quality went live, it took business analysts just four days to develop 20 Basel II business rules on the fly, deliver a data accuracy scorecard, create profiles on account history tables, and develop other business rules that were then added to the scorecard. "Basel II has been a blessing in disguise. Improved data quality and the ability to measure that quality have impacted our business for the better in several areas," said Hershaw. "It is clear that business analysis has benefited, but also there are factors such as how we evaluate our sales and marketing. We are now able to ascertain with confidence what activities have caused a sale.

"We live by the mantra that customers must be presented with no surprises, but we also experienced no surprises of our own in meeting the requirements of Basel II. It was a tough assignment, but the detailed planning and execution enabled us to minimize the scale of the challenge. We're now even more confident that we will always deliver on our commitments to customers and can assure high levels of data quality to assist in running our business," Hershaw concluded. ●

LESSON FROM THE EXPERTS

Data Quality Monitoring: The Basis for Ongoing Information Quality Management

By Tom Golden

Director of Corporate Communications (Data Quality), Informatica Corporation

The key elements of a good data quality program include establishing a baseline, continuous improvement, appropriate metrics, and scorecarding.

Establishing a Baseline

The first step is establishing a baseline of the current state of data quality. This should identify the critical failure points and determine improvement targets. The targets must be tied to business objectives.

Continuous Measurement

Data quality must be tracked, managed, and monitored if it is to improve business efficiency and transparency. Therefore, being able to measure and monitor data quality throughout the lifecycle and compare the results over time is an essential ingredient in the proactive management of ongoing data quality improvement and data governance.

Organizations need a formalized way of setting targets, measuring conformance to those targets, and effectively communicating tangible data quality metrics to senior management and data owners. Standard metrics provide everyone (executives, IT, and line-of-business managers) with a unified view of data and data quality, and can also provide the basis for regulatory reporting in certain circumstances, such as Basel II, where there are specific data quality reporting requirements.

Metrics to Suit the Job

Ultimately, data quality monitoring and reporting based on a well-understood set of metrics provides important knowledge about the value of the data in use, and empowers knowledge workers with the ability to

determine how the data can best be used to meet their own business needs.

The critical attributes of data quality (completeness, conformity, consistency, accuracy, duplication, and integrity) should map to specific business requirements. Duplicate records in a data warehouse, for example, make it difficult to analyze customer habits and segment customers in terms of market. Inaccurate data results in poor targeting, budgeting, staffing, unreliable financial projections, and so on. (The Informatica white paper, "Monitoring Data Quality Performance Using Data Quality Metrics," outlines a more comprehensive list of metrics and examples.)

A well-defined set of metrics should be used to get a baseline understanding of the levels of data quality; this baseline should be used to build a business case to justify investment in data quality. Beyond that, the same metrics become central to the ongoing data quality process, enabling business users and data stewards to track progress and quickly identify problem areas that need to be addressed.

Breaking down data issues into these key measures highlights where best to focus your data quality improvement efforts by identifying the most important data quality issues and attributes based on the lifecycle stage of your different projects. For example, early in a data migration, the focus may be on completeness of key master data fields, whereas the implementation of an e-banking system may require greater concern with accuracy during individual authentication.

Scorecarding

Inherent in the metrics-driven approach is the ability to aggregate company-wide results into data quality scorecards. A scorecard is the key visual aid that helps to drive the

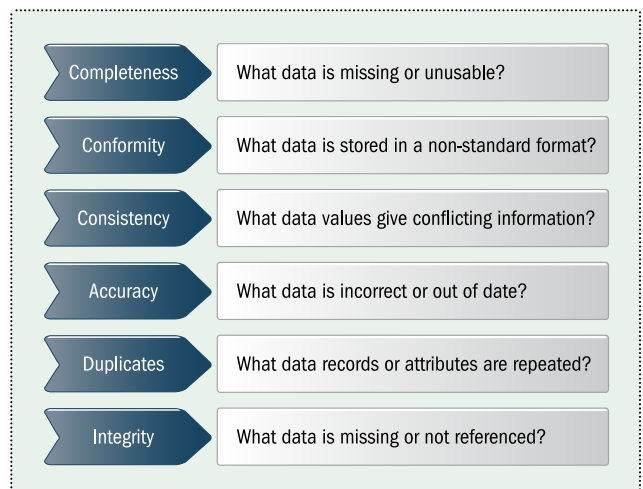


Figure 1. The critical attributes of data quality should map to specific business requirements.

data quality process in the right direction, empowering data analysts to set accurate and focused quality targets and to define improvement processes accordingly, including setting priorities for data quality improvement in upstream information systems.

Metrics and scorecards that report on data quality, audited and monitored at multiple points across the enterprise, help to ensure data quality is managed in accordance with real business requirements. They provide both the carrot and the stick to support ownership, responsibility, and accountability. But, beyond the data quality function itself, the metrics used for monitoring the quality of data can actually roll up into higher-level performance indicators for the business as a whole. ●

For free white papers on this topic, [download](#) "Profiling: Calculating Return on Investment for Data Migration and Data Integration Projects" or "Monitoring Data Quality Performance Using Data Quality Metrics," or [click here](#) for more information about Informatica Corporation.

Predictive Analytics

EXTENDING THE VALUE OF YOUR DATA WAREHOUSING INVESTMENT

BY WAYNE W. ECKERSON

WHAT IS PREDICTIVE ANALYTICS?

Predictive analytics can help companies optimize existing processes, better understand customer behavior, identify unexpected opportunities, and anticipate problems before they happen. Almost all of TDWI’s Leadership Award winners in the past six years have applied predictive analytics in some form or another to achieve breakthrough business results.

HIGH VALUE, LOW PENETRATION. With such stellar credentials, the perplexing thing about predictive analytics is why so many organizations have yet to employ it. According to our research, only 21% of organizations have “fully” or “partially” implemented predictive analytics, while 19% have a project “under development” and a whopping 61% are still “exploring” the issue or have “no plans.” (See Figure 1.)

Predictive analytics is also an arcane set of techniques and technologies that bewilder many business and IT managers. It stirs together statistics, advanced mathematics, and artificial intelligence and adds a heavy dose of data management to create a potent brew that many would rather not drink! They don’t know if predictive analytics is a legitimate business endeavor or an ivory tower science experiment run wild.

WHERE DO YOU START?

But once managers overcome their initial trepidation, they encounter another obstacle: how to apply predictive analytics optimally in their company. Most have only a vague notion about the business areas or applications that can benefit from predictive analytics.

STATUS OF PREDICTIVE ANALYTICS

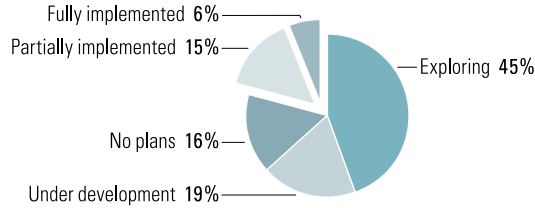


Figure 1. Predictive analytics is still in early-adopter phase. Based on 833 respondents to a TDWI survey conducted in August 2006.

Second, most don’t know how to get started: whom to hire, how to organize the project, or how to architect the environment.

DEFINITIONS

Before we address those questions, it’s important to define what predictive analytics is and is not. Predictive analytics is a set of business intelligence (BI) technologies that uncovers relationships and patterns within large volumes of data that can be used to predict behavior and events. Unlike other BI technologies, predictive analytics is forward-looking, using past events to anticipate the future. (See Figure 2.)

THE SPECTRUM OF BI TECHNOLOGIES

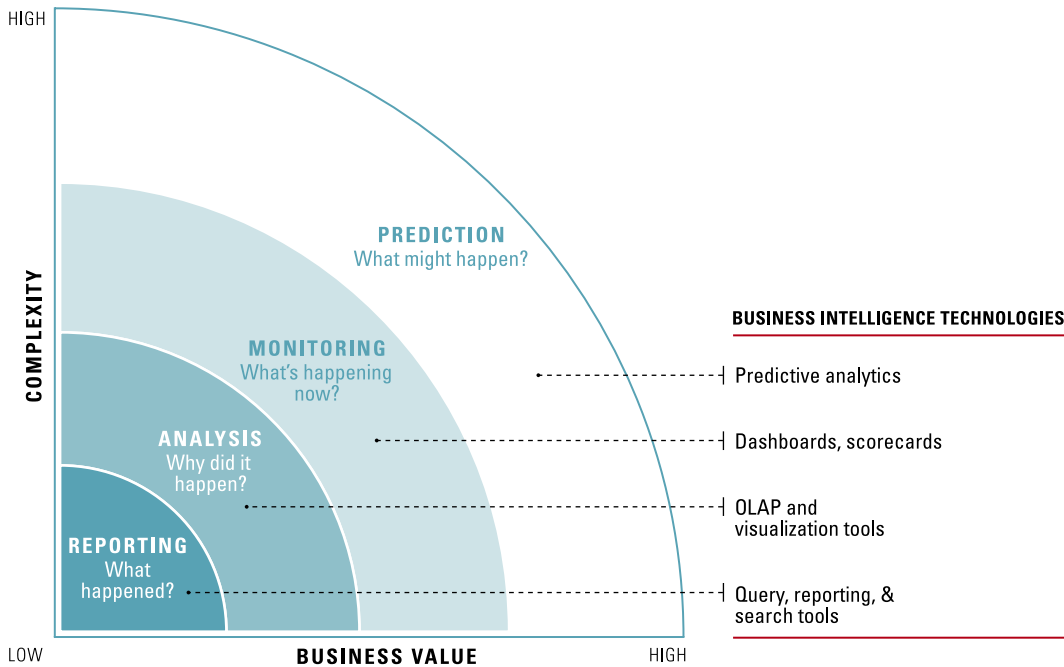


Figure 2. Among business intelligence disciplines, prediction provides the most business value but is also the most complex. Each discipline builds on the one below it—these are additive, not exclusive, in practice.

APPLICATIONS. Predictive analytics can identify the customers most likely to churn next month or to respond to next week’s direct mail piece. It can also anticipate when factory floor machines are likely to break down or figure out which customers are likely to default on a bank loan. Today, marketing is the biggest user of predictive analytics with cross-selling, campaign management, customer acquisition, and budgeting and forecasting models top of the list, followed by attrition and loyalty applications. (See Figure 3.)

APPLICATIONS FOR PREDICTIVE ANALYTICS

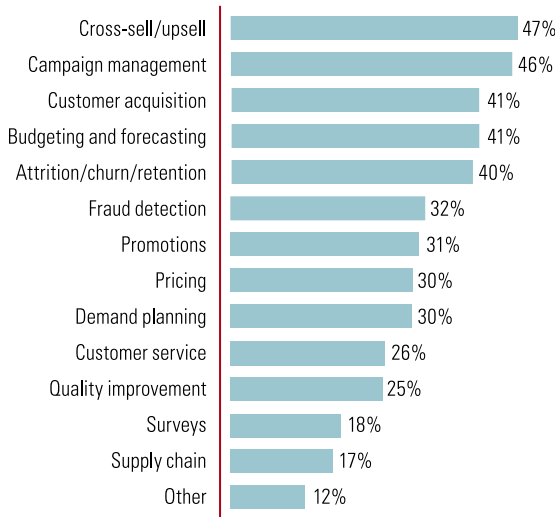


Figure 3. Based on 167 respondents who have implemented predictive analytics. Respondents could select multiple answers.

VERSUS BI TOOLS. In contrast, other BI technologies—such as query and reporting tools, online analytical processing (OLAP), dashboards, and scorecards—examine what happened in the past. They are *deductive* in nature—that is, business users must have some sense of the patterns and relationships that exist within the data based on their personal experience. They use query, reporting, and OLAP tools to explore the data and validate their hypotheses. Dashboards and scorecards take deductive reasoning a step further: they present users with a de facto set of hypotheses in the form of metrics and key performance indicators (KPIs) that users examine on a regular basis.

Predictive analytics works the opposite way: it is *inductive*. It doesn’t presume anything about the data. Rather, predictive analytics lets data lead the way. Predictive analytics employs statistics, machine learning, neural computing, robotics, computational mathematics, and artificial intelligence techniques to explore *all* the data, instead of a narrow subset of it, to ferret out meaningful relationships and patterns. Predictive analytics is like an “intelligent” robot that rummages through all your data until it finds something interesting to show you.

NO SILVER BULLET. However, it’s important to note that predictive analytics is not a silver bullet. Practitioners have learned that most of the “intelligence” in these so-called decision automation systems comes from humans who have a deep understanding of the business and know where to point the tools, how to prepare the data, and how to interpret the results. Creating predictive models requires hard work, and the results are not guaranteed to provide any business value. For example, a model may predict that 75% of potential buyers of a new product are male, but if 75% of your existing customers are male, then this prediction doesn’t help the business. A marketing program targeting male shoppers will not yield any additional value or lift over a more generalized marketing program.

THE BUSINESS VALUE OF PREDICTIVE ANALYTICS

INCREMENTAL IMPROVEMENT. Although organizations occasionally make multi-million dollar discoveries using predictive analytics, these cases are the exception rather than the rule. Organizations that approach predictive analytics with a “strike-it-rich” mentality will likely become frustrated and give up before reaping any rewards. The reality is that predictive analytics provides incremental improvement to existing business processes, not million-dollar discoveries.

“We achieve success in little percentages,” says a technical lead for a predictive analytics team in a major telecommunications firm. She convinced her company several years ago to begin building predictive models to identify customers who might cancel their wireless phone service. “Our models have contributed to lowering our churn rate, giving us a competitive advantage.”

The company’s churn models expose insights about customer behavior that the business uses to improve marketing or reengineer business processes. For example, salespeople use model output to make special offers to customers at risk of churning, and the managers to change licensing policies that may be affecting churn rates.

MEASURING VALUE. Our survey reinforces the business value of predictive analytics. Among respondents who have implemented predictive analytics, two-thirds (66%) say it provides “very high” or “high” business value. A quarter (27%) claim it provides moderate value and only 4% admit it provides “low” or “very low” value. (See Figure 4.)

Respondents who selected “very high” or “high” in Figure 4 say they measure the success of their predictive analytics efforts with several criteria, starting with “meets business goals” (mentioned by 57% of respondents). Other success criteria include “model accuracy” (56%), “ROI” (40%), “lift” (35%), and “adoption rate by business users” (34%). (See Figure 5.)

WHAT IS THE BUSINESS VALUE OF PREDICTIVE ANALYTICS TO YOUR ORGANIZATION?

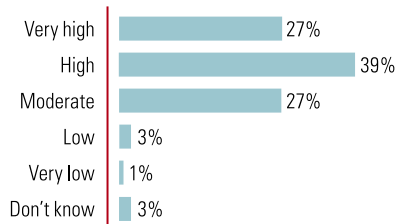


Figure 4. Based on 166 respondents who have implemented predictive analytics.

HOW DO YOU MEASURE SUCCESS?

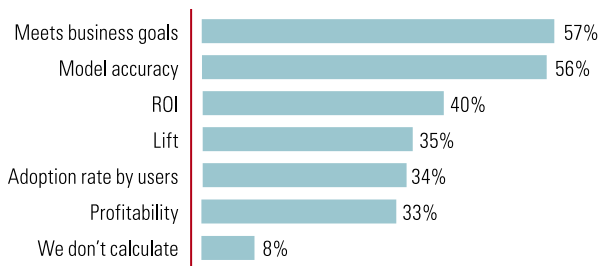


Figure 5. Based on 110 users who have implemented predictive analytics initiatives that offer “very high” or “high” value. Respondents could select multiple choices.

HOW DO YOU DELIVER PREDICTIVE ANALYTICS?

WHAT NOW? While some organizations have discovered the power of predictive analytics to reduce costs, increase revenues, and optimize business processes, the vast majority are still looking to get in the game. Today, most IT managers and some business managers understand the value that predictive analytics can bring, but most are perplexed about where to begin.

“We are sitting on a mountain of gold but we’re not mining it as effectively as we could,” says Michael Masciandaro, director of business intelligence at Rohm & Haas, a global specialty materials manufacturer. “We say we do analytics, but it’s really just reporting and OLAP.”

Rohm & Haas has hired consultants before to build pricing models that analyze and solve specific problems, but these models lose their usefulness once the consultants leave. Masciandaro says building an internal predictive analytics capability could yield tremendous insights and improve the profitability of key business areas, but he struggles to understand how to make this happen.

“How do you implement advanced analytics so they are not a one-off project done by an outside consultancy?” asks Masciandaro. “How do you bring this functionality in-house and use it to deliver value every day? And where do you find people who can do this? There are not too many of them out there.”

THE PROCESS OF PREDICTIVE MODELING

METHODOLOGIES. Although most experts agree that predictive analytics requires great skill—and some go so far as to suggest that there is an artistic and highly creative side to creating models—most would never venture forth without a clear methodology to guide their work. In fact, process is so important that in 1996 several industry players created an industry standard methodology called the Cross Industry Standard Process for Data Mining (CRISP-DM).

Regardless of methodology, most processes for creating predictive models incorporate the following steps:

- 1. PROJECT DEFINITION:** Define the business objectives and desired outcomes for the project and translate them into predictive analytic objectives and tasks.
- 2. EXPLORATION:** Analyze source data to determine the most appropriate data and model building approach, and scope the effort.
- 3. DATA PREPARATION:** Select, extract, and transform data upon which to create models.
- 4. MODEL BUILDING:** Create, test, and validate models, and evaluate whether they will meet project metrics and goals.
- 5. DEPLOYMENT:** Apply model results to business decisions or processes. This ranges from sharing insights with business users to embedding models into applications to automate decisions and business processes.
- 6. MODEL MANAGEMENT:** Manage models to improve performance (i.e., accuracy), control access, promote reuse, standardize toolsets, and minimize redundant activities.

Most experts say the data preparation phase of creating predictive models is the most time-consuming part of the process, and our survey data agrees. On average, preparing the data occupies 25% of total project time. However, model creation, testing, and validation (23%) and data exploration (18%) are not far behind in the amount of project time they consume. This suggests that data preparation is no longer the obstacle it once was.

RECOMMENDATIONS

Now that we’ve defined predictive analytics, assessed its business value, and stepped through key trends and processes, it’s important to provide specific recommendations to BI managers and business sponsors about how to implement a predictive analytics practice. This section offers five recommendations that synthesize best practices from various organizations that have implemented predictive analytics.

1. HIRE BUSINESS-SAVVY ANALYSTS TO CREATE MODELS

Every interviewee for this report said the key to implementing a successful predictive analytics practice is to hire analytic modelers with deep understanding of the business, its core processes, and the data that drives those processes.

“Good analysts need strong knowledge of narrow business processes,” says Keith Higdon of Sedgwick CMS. In claims processing, Higdon says good analysts “understand the business of claims handling; the interplay of variables across claim, claimant, and program characteristics; and what data they can and cannot rely on.” Only then, he adds, can analysts create “meaningful models” that result in positive business outcomes.

2. NURTURE A REWARDING ENVIRONMENT TO RETAIN ANALYTIC MODELERS

Since good analytic modelers are difficult to find, it’s imperative that managers create a challenging and rewarding environment. Of course, money is requisite. Top-flight analytic modelers often command higher salaries than classic business or data analysts. But good analytic modelers are motivated by things other than money and status, says Tom Breur, principal of XLNT Consulting in the Netherlands.

“You don’t attract analytic modelers with the same incentives as other people,” he says. “They want an opportunity to demonstrate their skills and learn new things, so you have to increase their training budgets. I’ve struggled with human resources departments on this issue.”

3. FOLD PREDICTIVE ANALYTICS INTO THE INFORMATION MANAGEMENT TEAM

THE INSIDE TRACK. Traditionally, analytics teams are sequestered away in a back room somewhere and report to an ambitious department head (usually sales or marketing) who is seeking to ratchet up sales and get an edge on the competition. Unfortunately, this approach is not optimal, according to most practitioners. Analytic modelers are voracious consumers of data, and must establish strong alliances with the data warehousing team to maintain access to the data.

“Since I work in the data warehousing department within IT, I go to my colleagues and say, ‘I need this,’ and I usually get it. I get to the head of the queue faster than someone from outside, which means I get more [storage] space, quicker access to data, and more privileges. As a result, my projects get pushed faster. Those are the unwritten rules,” says an analytic modeler.

4. LEVERAGE THE DATA WAREHOUSE TO PREPARE AND SCORE THE DATA

Once you’ve hired the people and established the organization, the next important task is to provide a comprehensive

solution for managing the data that the analytic modelers may want to use. While it is not necessary to build a data warehouse to support the analytic process, a data warehouse can make the process infinitely easier and faster.

SAVING TIME. A data warehouse pulls together all relevant information about one or more domains (e.g., customers, products, suppliers) from multiple operational systems and then integrates and standardizes this information so it can be queried and analyzed. With a data warehouse, analysts only have to query one source instead of multiple sources to get the data they need to build models.

5. BUILD AWARENESS AND CONFIDENCE IN THE TECHNOLOGY

One of the toughest challenges in implementing analytics is convincing the business that this mathematical wizardry actually works. “Building confidence is a big challenge,” says one analytics manager. “It takes a while before business people become confident enough in the models to apply the results. The ultimate litmus test is when business people are willing to embed models within operational processes and systems. That’s when analytics goes mainstream in an organization.”

But getting to a lights-out analytical environment is not easy. Most organizations, even those with large analytic staffs, still only apply analytics in pockets and have yet to truly operationalize the results.

CONCLUSION

Applying these five recommendations should enable any organization to implement predictive analytics with a good measure of success. While many people seem intimidated by predictive analytics because of its use of advanced mathematics and statistics, the technology and tools available today make it feasible for most organizations to reap value from predictive analytics. ●

Wayne W. Eckerson is the director of research and services for TDWI and author of *Performance Dashboards: Measuring, Monitoring, and Managing Your Business* (John Wiley & Sons, 2005). He can be reached at weckerson@tdwi.org.

This article was excerpted from the full, 32-page report by the same name. You can download this and other TDWI Research free of charge at www.tdwi.org/research.

The report was sponsored by MicroStrategy, Inc., OutlookSoft Corporation, SAS, SPSS, Sybase, Inc., and Teradata, a division of NCR.

CASE STUDY

Data Warehouse Appliance Combines Unconstrained Analytics with Energy Efficiency for UK's Leading Mobile Carrier

Commentary by Stephen Hawkins
Senior Data Architect, Orange UK

Customer Profile

Orange UK is the United Kingdom's most popular mobile phone service, with more than 13 million active subscribers. Orange UK is at the forefront of efforts by mobile communications carriers to develop sophisticated marketing programs that deliver targeted products and services to customers.

Business Challenge

Orange's success and future growth depends on in-depth insight into customer calling patterns and recognition of behavior trends through the analysis of call detail records (CDRs). As data volume continued to increase, Orange determined that it needed a new data warehousing platform. Orange needed to be able to analyze billions of CDRs in a fraction of the time used by other systems, while reducing its equipment footprint.

Deeper, Wider, Faster

With its Intelligent Query Streaming architecture, the Netezza Performance Server (NPS) data warehouse appliance has allowed Orange to broaden and deepen its BI efforts while using its existing analytical tools. Orange is now able to run more queries than it could before, with faster query times. As a result, Orange's BI team has been able to dramatically increase its responsiveness to internal business users. Orange can now analyze billions of CDRs to offer highly targeted promotions, while improving the ability to detect fraud and capture cross-carrier billing.

The system is also able to keep pace with rapidly rising data volumes. "Since we first installed the Netezza system, we have added

at least half as much data again to the database, yet still achieve the same speed of response," explains Sylvia Sawkins, manager of data control, Orange UK.

Orange's BI team has been able to dramatically increase its responsiveness to internal business users.

Orange can now analyze billions of CDRs to offer highly targeted promotions, while improving the ability to detect fraud and capture cross-carrier billing.

Substantial Space and Power Savings

With space and power at a premium within Orange's data centers, the NPS system saved large amounts of both. With its previous system, Orange was consuming almost 25,000 watts and 85,000 BTU per hour. Once it implemented a three-terabyte Netezza data warehouse appliance, Orange saw its power consumption drop to 7,000 watts, and cooling requirements were reduced to 24,000 BTU per hour.

Even as the amount of information continues to grow, Netezza's compact data warehouse appliance has allowed Orange to significantly reduce the equipment footprint in its data center, with the number of cabinet spaces dropping from 26 to 9. The speed and ease of deployment were an added bonus.

Market Leader Orange Raises the BI Bar

For Orange UK, Netezza's data warehouse appliance has significantly improved data mining on CDRs, enabling highly targeted marketing to ensure that the highest value is obtained from each customer, while maximizing retention. And by reducing complexity, power consumption, and space requirements—three of the biggest costs facing IT departments—the Netezza system provides an efficient platform for future growth.

Now that the NPS appliance enables timely access to terabytes of data, Orange is able to turn its attention to new business challenges. "The system started by making our lives easier; this led to an ability to handle a greater capacity. Culturally, now the business demands more from its data, expectations are set much higher, so there's a great focus on quality," Sawkins adds. The goal: data that is not just faster, but better. ●

LESSON FROM THE EXPERTS

Data Warehouse Appliances Help Tackle the Data Center Power Crisis

By **Phil Francisco**

Director of Product Marketing, Netezza Corporation

Increasing power costs and increasingly constrained power availability are prompting many companies to reexamine their IT strategies and budgets. Companies are facing a power dilemma—do they scale back growth plans, pay the cost of building an additional facility, or relocate the data center where costs are lower?

Smaller... Faster... Hotter

Data center servers used for general-purpose computing are dominated by Intel/AMD, or x86, architectures. This chip technology is at the heart of blade servers and rack servers—including those used in other data warehouse systems—and contains features designed to improve performance for general-purpose computing. But the increasing data volumes and performance demands of data warehousing are putting the CPU under tremendous strain. CPU designs based on x86 architectures are moving more data through smaller components at faster speeds than ever before. With so much processing activity and input/output traffic condensed into a very small space, the devices produce much more heat, and require significantly more cooling, than their predecessors of only a few years ago.

Green and Greener

The industry is responding to this in several ways. Chip manufacturers are announcing CPUs that deliver higher performance than previous top-of-the-line chips, yet draw less power. Industry consultants and experts abound with advice on how best to cool the data center or improve its the energy efficiency.

But for data warehousing, greater energy and space efficiency has already arrived in the form of a data warehouse appliance—the

Netezza Performance Server (NPS) system. The NPS system is built specifically for high-performance data warehousing, while consuming significantly less power and generating less heat than other solutions. Netezza's architecture uses intelligent storage nodes to process data as it streams off the disk, increasing performance. Each of these nodes uses an embedded PowerPC chip that consumes 4.5 watts of power (compared to 70 or more watts in x86-based systems), reducing the power typically required by other data warehouse solutions. The architecture of Netezza's data warehouse appliance—rather than expensive components—makes the difference.

Customers process massive queries in significantly shorter periods of time, with much lower electricity consumption and minimal space requirements that translate into lower cost of ownership.

Low Power, High Performance

Netezza's architecture takes a different approach to processing queries than architectures developed for general-purpose computing. This not only provides performance gains; it also lowers power consumption and heat.

Unlike many traditional data warehousing systems, Netezza's asymmetric massively parallel processing (AMPP) architecture is built for streaming processing of data. The system's architectural approach uses commodity field programmable gate arrays

(FPGAs) to do the bulk of the data filtering, with an embedded PowerPC chip handling the remainder. The AMPP architecture, rather than expensive components, provides the performance difference—along with substantial energy savings.

The Energy-Efficient Data Warehouse

The Netezza data warehouse appliance, with streaming processing on its energy-efficient intelligent storage nodes, consumes a fraction of the power required by systems based on other leading processors and general computing architectures. Customers process massive queries in significantly shorter periods of time, with much lower electricity consumption and minimal space requirements that translate into lower cost of ownership. ●

For a free white paper on this topic, [download](#) "Large Scale Use of the Netezza Performance Server System," or [click here](#) for more information about Netezza Corporation.

NETEZZA
The Power to Question Everything™

LESSON FROM THE EXPERTS

HP-Oracle Reference Configurations for Data Warehousing

By **Chris Buss**, World Wide Manager, BCS BI Marketing, HP
 By **Bill Nee**, Senior Director, Database Product Marketing, Oracle Corporation

HP and Oracle jointly offer a rich set of reference configurations that provide standard components, predictable performance and rapid deployment for data warehouses between 200 GB and 1 TB in size. These reference configurations are built on the combined knowledge and experience of these two industry leaders and are an important component of the Oracle Information Appliance Initiative.

Challenges

IT planners and architects have a wide range of technology options for their data warehouse (DW) applications. Each hardware platform decision requires them to research, and then balance, the tradeoffs between price, performance, long-term scalability, manageability, and vendor reputation. When rapid deployment is a key factor, the choice often spans dedicated DW hardware, rigidly prescribed single-vendor configurations, and building custom solutions from well-understood technologies. Such a range in choices often means complex, time-consuming decision cycles that drain staff time by necessitating the definition and management of benchmarks or proofs of concept. All of this takes away from the basic function of IT, which is to deliver information quickly to business users so that they can make sound, timely decisions.

Solution

As reported by the leading analyst firm IDC, HP is the world leader in infrastructure for strategic DW, and Oracle is the leading DW database. Together, HP and Oracle support many of the world's largest data warehouse deployments, as well as tens of thousands of smaller implementations.

To reduce the risk and deployment time for DW solutions, HP and Oracle have created

a comprehensive set of reference configurations for Oracle's Information Appliance Initiative. These configurations draw upon both companies' extensive lab and performance benchmark investments as well as global experience supporting thousands of data warehouse and business intelligence customers. Reference configurations

allow IT planners and architects to start from a "best-fit" hardware configuration that reflects their raw data size, database, operating system, processor, and node scaling preference. These configurations can then be optimized based on a customer's specific workload and requirements. For convenience, they are arranged into size groups, each scoped for a specific range of raw data sizes. Within each group, HP and Oracle offer a rich set of architecture and performance options.

Each configuration reflects real-world DW customer environments and workloads, balancing processing power, storage, and throughput to provide predictable and cost-effective high performance. Achieving balanced system input/output (I/O) throughput specifically designed to satisfy the requirements of DW workloads was a key objective in developing the configurations. And HP-Oracle reference configurations are modular and can easily scale up by adding CPUs, nodes, or storage on an as-needed basis.

Some vendors encourage customers to purchase dedicated, proprietary hardware that cannot be repurposed as business needs change. Others recommend solutions from a narrowly prescribed set of single-vendor components. Neither approach delivers on the promise of open, flexible, IT solutions based on standardized technologies. Oracle-HP reference configurations combine architectural choice with predictable performance and standards-based technology.

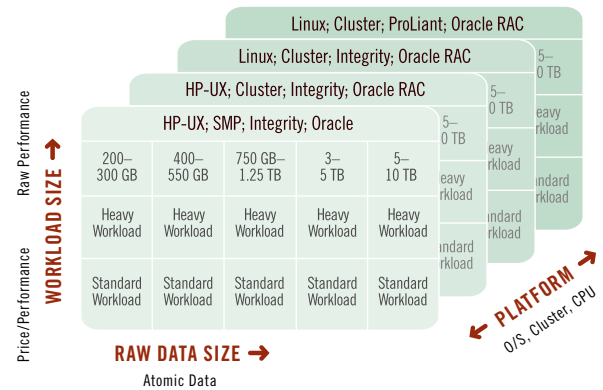


Figure 1. HP-Oracle reference configurations provide 40 configurations in a matrix model that covers key DW design points.

Benefits

IT staff time and resources are precious. HP-Oracle reference configurations enable IT departments to do more with less, more quickly, and at lower risk. Some immediate benefits for deploying data warehouse solutions built using these Reference Configurations include:

- Simplified product selection and sizing process
- A predefined set of core components to simplify purchasing
- Simplified, rapid deployment with predictable performance
- Off-the-shelf components to enable repurposing and lower risk
- Proven, award-winning support, with access to performance and tuning services

In summary, HP-Oracle reference configurations provide optimal DW performance and enable customers to avoid common design pitfalls. Customers can more quickly and easily deploy high-performance data warehouse solutions for accelerated time-to-intelligence, reduced costs, and increased revenue. ●

For more information, visit:

- <http://www.oracle.com/features/hp/data-warehousing.html>
- www.hp.com/go/bi

CASE STUDY

Aquascutum Innovates in Retail

Commentary by Robin Coles

Business Development Manager, Prologic

Customer Overview

Aquascutum was founded in 1851 by a tailor in the heart of Mayfair, London. Aquascutum's early business was focused on developing an innovative new attribute for its coats: the "shower-proofing" of wools. Today, Aquascutum is a leading fashion house that offers designer clothing for men and women, luggage, and accessories.

Business Challenge

Aquascutum relies on CIMS, a dedicated retail management application, from Prologic to manage key areas of the business, including merchandising, warehousing, distribution, allocation, replenishment, and sourcing. CIMS is an end-to-end enterprise resource planning (ERP) solution that has been deployed at more than 35 fashion and lifestyle companies in the U.K.

In the dynamic world of fashion, retailers such as Aquascutum need timely access to detailed business information, enabling them to react to trends while managing inventory and making a host of other critical business decisions. Aquascutum uses Prologic's sophisticated reporting tool—CIMS Vision—to deliver the information required within its dynamic business. However, as the data volumes and requirements have grown, the company began to find that the system lacked the back-end power required to deliver timely information. As a result, a number of reports slowed to the point where basic information could take several hours to generate.

In addition, Aquascutum was anxious to better exploit the information CIMS Vision provided by generating ad hoc and "what if" analyses; this had proved impractical due to the sheer volume of data contained within its data warehouse.

The Solution

To meet this need, Prologic integrated HyperRoll into the Aquascutum CIMS Vision application. Aquascutum can now provide its users with lightning-fast reporting capabilities and enable users to perform increasingly sophisticated queries. HyperRoll combines patented query acceleration, caching, and compression algorithms to deliver immediate and substantial improvements in business intelligence performance. HyperRoll's technology works seamlessly alongside CIMS Vision to accelerate data loading by up to factors of 200, while enabling access and analysis of data in unparalleled detail.

.....

Enormous amounts of data can now be summarized at the touch of a button, improving business visibility and enabling users to become more proactive in their approach to merchandising.

Robin Coles, Prologic

.....

"Using HyperRoll's data aggregation software, we are now able to offer substantially increased analysis capabilities to our customers. Enormous amounts of data can now be summarized at the touch of a button, improving business visibility and enabling users to become more proactive in their approach to merchandising," said Robin Coles, business development manager of Prologic. "We anticipate wide adoption of the HyperRoll-enabled CIMS Vision solution, both within our customers' organizations, and the market as a whole."

The Results

By integrating HyperRoll into the CIMS Vision reporting environment, Aquascutum has dramatically improved its query performance

and thus empowered analysts and decision-makers to achieve rapid business insight. For example, a weekly business key performance indicator (KPI) report that used to take 80 minutes to open and up to 30 minutes to print, now takes only 2 minutes to open and 30 seconds to print.

"If our queries take hours to run, it is very difficult to analyze our business and make strategic decisions fast enough to make a difference," said Craig Kayes, business projects manager of Aquascutum. "Prologic and HyperRoll have allowed us to analyze our key business drivers on an ongoing basis and proactively use our data in our decision-making process. We anticipate that these improved reporting capabilities will continue to have a dramatic impact on our business."

Such reductions in query response times are already providing benefits to multiple areas of the business. Aquascutum is now able to report in detail on every single style in its business, perform ad hoc queries, and conduct extensive "what-if" analyses to better inform a range of business decisions. The company is now able to respond to situations more quickly, while identifying issues and creating strategies to address current opportunities or threats. In addition, they are able to drill down into detail, identifying the root causes and rapidly resolving them. ●

LESSON FROM THE EXPERTS

Investing in Data Warehouse Query Performance

By **Eric Rogge**

Vice President of Marketing, HyperRoll

There are many challenges with data warehouses, not the least of which is estimating the investment in software, hardware, and effort to optimize the performance of a data warehouse. It is important that data warehouse architects, managers, and other stakeholders enter into the data warehouse investment cycle with honest and realistic assessments of performance need. Of course, the best plans can go awry. Plus, there is always a political aspect to the investment, where commitments are made and expectations are set by vendors, by external consultants, by internal architects, and by IT managers. Any unmet expectation by the data warehouse or by these stakeholders can damage business relationships and careers.

With all of these factors in play, defining the level of investment needed for adequate data warehouse performance can be a slippery fish. To capture the fish, or in other words, to identify adequate investment, key stakeholders should itemize and plan for the different data warehouse capabilities that are impacted by its performance. These include:

- **Static to dynamic report transition**

Report development backlogs are often addressed by converting static report sets to smaller, more versatile prompted or ad hoc reporting capabilities to users. The engineering trade-off is that user query rates and complexity becomes less predictable and manageable.

- **Query complexity**

To compete in the market, businesses look to increasingly complex analytics to manage the subtle and complex aspects of competing for customer attention and engagement. These analytics drive increasingly complex and performance-demanding queries to the data warehouse.

- **Aggregates**

Most queries to a data warehouse are aggregate queries, requiring extra computational work by the relational database management system (RDBMS). Supporting a few aggregates is easy, but adding dimensional permutations can quickly scale the number of aggregates to calculate to thousands or more.

- **Schema complexity**

As with query complexity, schemas are now more complex because they must describe more complex business operations. This is especially true as users demand new and different views of the data from the data warehouse.

- **User scalability**

Often business expects a nonlinear scaling of user communities. Early user counts may be in the low teens or hundreds, but full-scale deployments may be in the thousands.

Clearly, data warehouse managers need to forecast the rate and nature of change for these and other similar factors that impact data warehouse performance. The most common approach is to size the data warehouse system for near-term needs and then hope that traditional data warehouse tuning solutions will cover any performance gaps in the foreseeable future. The unfortunate truth is that this approach quite often falls short of meeting the need for data warehouse query performance. Combining the impact of a few of the above performance factors can require a tenfold or even a hundredfold improvement in query performance. With this level of query performance risk, data warehouse stakeholders have to take a progressive approach to assuring adequate performance.

That approach requires a fundamental understanding of all possible data warehouse hardware and software architecture approaches for assuring performance. Much as successful fishermen often have secret and unique ways of catching fish based upon deep knowledge of fish behavior and environment, data warehouse stakeholders should cultivate broad understanding of performance enhancement options and use innovative ways to meet their data warehouse performance demands. This means moving beyond the mainstream methods and seeking alternatives that are not in use by business competitors.

An evaluation framework for assuring data warehouse query performance should start with a broad search for query performance enhancement alternatives. This should be followed by a comparative assessment of the various alternatives based upon an organization's performance impacting factors (like those described here) along with the other usual technology and vendor selection criteria. Organizations should evaluate their need to differentiate from their competitors via speed and ease of data access. As the need for differentiation grows in priority, so should the need for differentiating technology and methodologies. Most importantly, a structured business case should be created for any initiative to increase data warehouse query performance. ●

For a free white paper on this topic, [download](#) "Ensuring Data Warehouse Query Performance," or [click here](#) for more information about HyperRoll.

CASE STUDY

TEOCO Increases Performance and Customer Satisfaction

Commentary by John Devolites

Vice President and General Manager,
Communications and Entertainment Solutions,
TEOCO

About TEOCO

Founded in 1994, TEOCO (The Employee Owned Company) is a leading provider of strategic business solutions to large and small enterprises. TEOCO's Communications and Entertainment Solutions group is the premier provider of network cost and revenue management solutions to the communications industry. TEOCO combines the strength of its renowned software applications with in-depth auditing expertise to achieve unrivaled market penetration and industry leadership in the communications space.

Business Problem

As continued competitive pressures drove consolidation in the communications industry, TEOCO found many of their customers merging together to create larger enterprises with sharply increased transaction volumes. Originally, TEOCO's systems were designed to support up to \$1 billion in monthly billing transaction volume and even less for reporting. However, with increasing customer transaction volumes, the company was experiencing substantial run-time issues and large report issues that resulted in delayed customer response. While the transaction processing was still adequate, the reporting delays were significant and continually degrading.

"Just to give you a very simple overview: AT&T merged with SBC, raising the transactions from \$1 billion worth to \$2 billion worth per month. Now, with the addition of Bell South and Cingular, there's another \$600 to \$700 million of transactions per month, resulting in our application processing nearly \$3 billion in transactions per month, but our systems were designed only to handle things in the billion dollars with even less range for reporting," says John Devolites. To better

serve its customers and drive competitive advantage, TEOCO decided that an alternate data warehouse architecture and environment must be implemented.

The Solution

TEOCO did its research and chose to evaluate data warehouse appliances on three major components, the first of which was performance. Devolites says, "At the top end of the spectrum we wanted something that would run at least 40 to 100 times faster on the same types of queries that we were dealing with today on Oracle."

"We basically built up 5 terabytes of data to be able run our benchmarks against. Those benchmarks were the same that we were running on the Oracle side of the equation. With the DATAlegro appliance, across the board, we saw a 50–100 times performance improvement on the queries that we were running. What that means is that a query that was taking 20–22 hours to run, ran in minutes!"

John Devolites, TEOCO

TEOCO set out to find a vendor that could offer a flexible, scalable, yet cost-effective solution. Therefore, a proof of concept was essential to the company's ability to validate the reliability and performance of any data warehouse appliance. "We basically built



up 5 terabytes of data to be able to run our benchmarks against. Those benchmarks were the same that we were running on the Oracle side of the equation. With the DATAlegro appliance, across the board, we saw a 50 to 100 times performance improvement on the queries that we were running. What that means is that a query that was taking 20 to 22 hours to run, ran in minutes," comments Devolites.

Tangible Benefits

By implementing a data warehouse appliance, TEOCO is now able to mine data in innovative ways. Devolites says, "So this new technology allows us to mine this information in a way we've never seen it before. We can look at in-office and call routing patterns. We can look at fraud. We can look at different things that we have never before been able to audit for a communications carrier.

"For one of the carriers, we ran through a whole series of cell phone billing information and identified their top 100 users. Within the top hundred users, there were 10 with over 25,000 minutes per month. Because of the DATAlegro appliance, we were able to look at data in a way that we've never been able to before and determine that it was a cab company in Seattle abusing a service plan. Before DATAlegro, there would have been no other way to determine that information," says John Devolites. ●

LESSON FROM THE EXPERTS

Appliances—Data Mart or Enterprise Data Warehouse?

By **Stuart Frost**
CEO, DATAlegro, Inc.

Appliances are now becoming well established in the data warehousing market. 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 people would claim that appliances are currently ready to handle complex EDW, they are finding an interesting niche as an integral part of many EDW infrastructures.

DM and an 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. For example:

- Significant data growth due to:
 - New legislation (Sarbanes-Oxley, E.U. data retention laws, etc.)
 - Mergers and acquisitions
- 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
- 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

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 the business needs!

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.

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 now 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 like:

- Data cleansing

By offloading these tasks from the EDW to the appliance, companies are greatly reducing the need for an expensive EDW upgrade. In addition, the specialized nature and advanced technology of the appliance enables the above 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 at the same time 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 a lot of cost? ●

For a free white paper on this topic, [download](#) “The Next Generation of Data Warehouse Appliances,” or [click here](#) for more information about DATAlegro, Inc.

Solution Providers

The following solution providers have shared their data integration stories and successes, technology insights, and the lessons they have learned for *What Works in Data Integration*.



Business Objects

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

Business Objects is the world's leading BI software company, helping organizations gain better insight into their business, improve decision making, and optimize enterprise performance. The company's business intelligence (BI) platform, BusinessObjects™ XI, offers the industry's most advanced and complete platform for reporting, query and analysis, performance management, and 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.



Collaborative Consulting

70 Blanchard Road, Ste. 500
Burlington, MA 01803
877.376.9900
Fax: 781.565.2700
info@collaborativeconsulting.com
www.collaborativeconsulting.com

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



DataFlux

940 NW Cary Parkway, Ste. 201
Cary, NC 27513
919.447.3000
Fax: 919.447.3100
sales@dataflux.com

DATAlegro, Inc.

85 Enterprise, 2nd Floor
Aliso Viejo, CA 92656
949.330.7690
Fax: 949.330.7691
info@datalegro.com
www.datalegro.com

Hyperion Solutions Corporation

5450 Great America Parkway
Santa Clara, CA 95054
408.744.9500
Fax: 408.588.8500
www.hyperion.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 v3™ is the industry's most advanced data warehouse appliance utilizing an all-commodity platform. By combining DATAlegro's patent-pending software with the industry's leading hardware, storage, and database technologies, DATAlegro has taken data warehouse performance, reliability, and innovation to the next level. DATAlegro v3 goes beyond the low cost and high performance of first-generation data warehouse appliances and adds the flexibility and scalability that only a commoditized platform can offer.

Whether you have a few terabytes of user data or hundreds, DATAlegro's data warehouse appliances deliver a fast, flexible, and affordable solution that allows a company's data to grow at the pace of its business.

Hyperion Solutions Corporation (Nasdaq Global Select: HYSL) is the global leader in business performance management software. More than 12,000 customers in 90 countries rely on Hyperion both for insight into current business performance and to drive performance improvement. With Hyperion software, businesses collect, analyze, and share data across the organization, linking strategies to plans, and monitoring execution against goals. Hyperion integrates financial management applications with a business intelligence platform into a single management system for the global enterprise. And with Hyperion services, your organization can confidently implement, learn, and run your management system. To find out more, please visit www.hyperion.com.

HyperRoll

INFORMATICA®

The Data Integration Company™

Information Builders

HyperRoll

1300 Crittenden Lane, Ste. 104
Mountain View, CA 94043
650.962.2571
Fax: 650.962.2501
sales@hyperroll.com
www.hyperroll.com

Informatica Corporation

100 Cardinal Way
Redwood City, CA 94063
650.385.5000
Fax: 650.385.5500
www.informatica.com

Information Builders, Inc.

Two Penn Plaza
New York, NY 10121-2898
212.736.4433
Fax: 212.967.6406
askinfo@ibi.com
www.informationbuilders.com

HyperRoll software is a data warehouse query performance engine that dramatically accelerates the performance of your existing data warehouses and business intelligence applications while also reducing your hardware, storage, and manpower costs. With HyperRoll, query speed improvements between 10x and 100x are typical. HyperRoll combines patented aggregation, memory-based caching, and compression algorithms to provide lightning-fast access to large data volumes while drastically reducing the need for additional system resources. HyperRoll supports 7x24 operating environments, is 100 percent compatible with your existing database and reporting tools, and will dramatically improve information access for your business.

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 track record of success that extends back to 1993, 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.

Information Builders created the industry's most widely deployed business intelligence solution and is the leader in real-time operational reporting. The WebFOCUS business intelligence platform, the company's flagship product, has the architecture, integration, and simplicity to permeate every level of the extended enterprise. It is the most scalable, secure, and flexible solution in the market, and helps organizations build applications that have no barriers. Information Builders is the only vendor with its own complete integration solution.

iWay Software's ability to seamlessly integrate with WebFOCUS enables unrivaled information access.

Information Builders' award-winning technology provides quality software and superior services to more than 12,000 customers, including most of the *Fortune* 100 and U.S. federal government agencies. Headquartered in New York City with 90 offices worldwide, the company employs 1,600 people and has more than 350 business partners.



Initiate Systems, Inc.

200 West Madison, Ste. 2300
Chicago, IL 60606
800.992.0300, 312.759.5030
Fax: 312.759.5026

info@initiatesystems.com
www.initiatesystems.com

Initiate Systems is the leading provider of customer-centric master data management solutions for companies that want to create the most complete, real-time views of people, households, and organizations from data dispersed across multiple application systems and databases. Initiate Identity Hub™ software is more accurate, scalable, rapidly deployed, and widely used than any other solution. Many organizations in the financial services, government, healthcare, hospitality, and retail sectors are using Initiate™ software. Initiate Systems' proven experience makes it uniquely qualified to enable strategic initiatives that allow organizations to increase revenue and efficiency and reduce operating costs and risks.

MicroStrategy

1861 International Drive
McLean, VA 22102
703.848.8600
Fax: 703.848.8610

info@microstrategy.com
www.microstrategy.com

MicroStrategy is a 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 their critical data actionable—quickly, simply, and affordably. The Netezza Performance Server® (NPS) family of products delivers breakthrough performance, unmatched ease of deployment and operation, and innovative flexibility and scalability. By architecturally integrating database, server, and storage within a single appliance, the NPS system delivers significantly faster performance than traditional systems, at a much lower total cost of ownership. Customers who have recognized the benefits of Netezza's approach include Ahold, Amazon.com, Debenhams, the U.S. Department of Veterans Affairs, Neiman Marcus, and Orange UK.



The Oracle logo consists of the word "ORACLE" in a bold, red, sans-serif font, with a registered trademark symbol (®) to the upper right.

The Petris logo features the word "PETRIS" in a bold, green, sans-serif font, with a stylized graphic of three green vertical bars of varying heights to the left.

Nimaya Inc.

7900 Westpark Drive, Ste. T-300
 McLean, VA 22102
 703.462.1291
 Fax: 703.462.1292
inquiries@nimaya.com
www.nimaya.com

Together, Nimaya and IBM help companies realize higher levels of customer satisfaction at lower costs, turning customers into advocates and increasing retention and revenue. Nimaya's CustomerGrid, a service-oriented architecture (SOA) application running on IBM WebSphere, leverages data residing in systems across divisions, functions, and extended enterprises. By capitalizing on investments in existing applications, Nimaya and IBM deliver inexpensive, easy-to-use solutions, addressing corporate needs for becoming customer-centric by providing consistent and accurate data. Key to this is Nimaya's business rules engine, which delivers insight and generates proactive leads, alerts, and tasks through unified customer views and enterprise data mash-ups.

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.

Petris Technology, Inc.

1900 St. James Place, Ste. 700
 Houston, TX 77056
 +1 713 956-2165
 Fax: +1 713 956-2185
sales@petris.com
www.petris.com

Since 1994, Petris has been providing practical, real-world information technology solutions for its energy clients, leveraging its expertise and experience in drilling and wellbore data management and analysis, geospatial information systems, and technical services.

Through a unique vendor-neutral approach and expertise in service-oriented IT architecture, Petris offers advanced data management technology that seamlessly integrates data from diverse data stores. The resulting software tools are intuitive to use and transparent to operate. While competitors offer solutions to address pieces of the data management puzzle, only Petris has the services and solution framework that includes them all.



Pitney Bowes Group 1 Software

4200 Parliament Place, Ste. 600
Lanham, MD 20706-1844
888.413.6763
Fax: 301.731.0360
Info@g1.com
www.g1.com

Pitney Bowes Group 1 Software provides innovative software solutions that enable our clients to better understand and connect with their millions of customers, prospects, and partners. Group 1 helps over 3,000 organizations improve profitability, increase effectiveness, and strengthen customer relationships through consolidating, cleansing, and enriching corporate data, and generating personalized business documents for multi-channel delivery, customer care, and efficient business processing. Our comprehensive Customer Communication Management (CCM) solutions span from database to delivery, adding value to every aspect of communication and allowing clients to integrate intelligence throughout their mailstream.

SAP

3999 West Chester Pike
Newtown Square, PA 19073
800.872.1727
info@sap.com
www.sap.com

SAP is the world's leading provider of business software. More than 38,000 customers in more than 120 countries run SAP® applications—from distinct solutions addressing the needs of small and midsize enterprises to suite offerings for global organizations. Powered by the SAP NetWeaver® platform to drive innovation and enable business change, SAP software helps enterprises of all sizes around the world improve customer relationships, enhance partner collaboration, and create efficiencies across their supply chains and business operations. SAP solution portfolios support the unique business processes of more than 25 industries, including high-tech, retail, financial services, healthcare, and the public sector. With subsidiaries in more than 50 countries, the company is listed on several exchanges, including the Frankfurt stock exchange and NYSE, under the symbol SAP. Additional information is available at: www.sap.com/usa/platform/netweaver/components/mdm/index.epx.

Siperian, Inc.

1820 Gateway Drive, Ste. 109
San Mateo, CA 94404
650.571.2400
Fax: 650.350.2206
www.siperian.com

Siperian offers the most complete, integrated software platform for adaptive master data management (MDM). Siperian helps companies create and present real-time unified views of their customers, products, and suppliers to business users from distributed data sources for higher profitability, reduced operational costs, and improved compliance. Our award-winning solution, Siperian MDM Hub™, uniquely adapts to different architectural styles to meet companies' evolving business needs while delivering significantly lower total cost of ownership, faster time-to-value, and superior return on investment. Siperian business solutions for financial services, health and life sciences, high-tech, communications and media, and manufacturing industries improve customer relationship management, marketing, regulatory compliance, and order-to-cash processes.



Sybase

One Sybase Drive
Dublin, CA 94568
925.236.5000

www.sybase.com

Syncsort Incorporated

50 Tice Boulevard
Woodcliff Lake, NJ 07677
201.930.9700
Fax: 201.930.8282

info@syncsort.com

www.syncsort.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, providing open, cross-platform solutions that securely deliver information anytime, anywhere—enabling customers to create an information edge. Sybase addresses the need for business intelligence and data warehousing with world-class data modeling, ultra-high-speed business analytics, and comprehensive data integration technologies. Maximize the potential of your information assets and achieve better intelligence capabilities with Sybase IQ, PowerDesigner, and our Data Integration Suite. When businesses get serious about business intelligence and data warehousing, they get Sybase. For more information, visit:

www.sybase.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.

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

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.

WORLD CONFERENCES

www.tdwi.org/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.

SEMINAR SERIES

www.tdwi.org/seminars

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.

ONSITE COURSES

www.tdwi.org/onsite

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.

CERTIFIED BUSINESS INTELLIGENCE PROFESSIONAL (CBIP)

www.cbipro.com

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.

WEBINAR SERIES

www.tdwi.org/education/Webinars

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.

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.





**You've got
the skills.
You've got
the experience.**



Now prove it.

You've taken classes and backed up the coursework with real-world experience. Now document your expertise by getting your Certified Business Intelligence Professional (CBIP) credential. This exam-based certification program tests your knowledge, skills, and experience, providing you with the most comprehensive and credible certification available in the industry.

Learn how to become a
Certified Business Intelligence Professional.

Visit www.cbipro.com.



VOLUME 23

WhatWorks
IN DATA INTEGRATION

1201 Monster Road SW
Suite 250
Renton, WA 98057

425.277.9126
info@tdwi.org
www.tdwi.org

