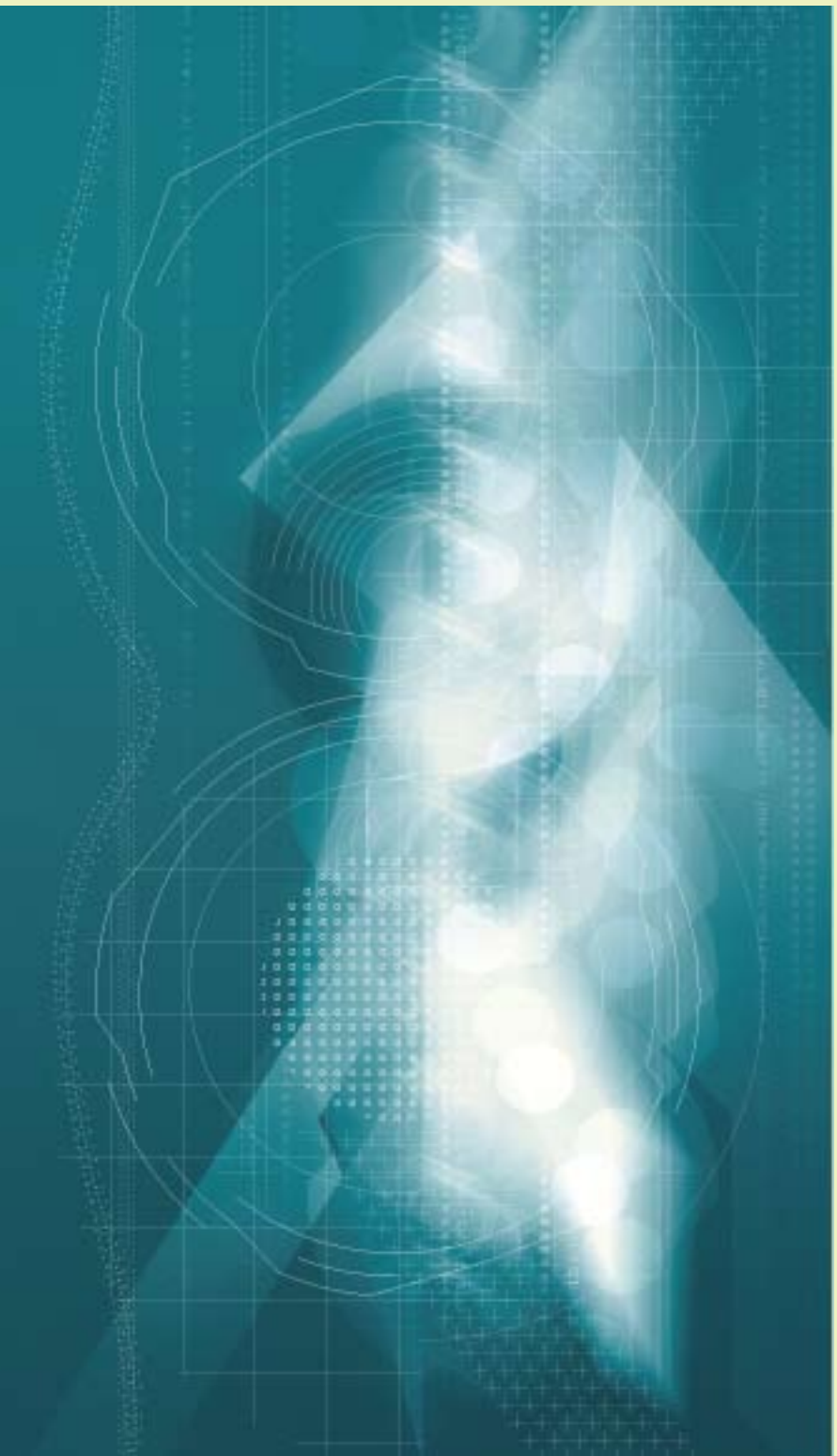
What Works

IN DATA INTEGRATION



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

Data Integration

Data Governance

Data Warehouse Appliances

Data Warehousing

Master Data Management

FEATURE

TDWI RESEARCH

tdwi
THE DATA WAREHOUSING INSTITUTE

Letter from the Editorial Director

This new edition of *What Works in Data Integration* offers a fresh, topically focused collection of customer success stories and expert perspectives. We're proud to offer this resource to enhance your understanding of the tools, technologies, and methods that are central to data integration today. We've arranged these case studies and lessons from the experts into specific categories to guide you through the articles: data integration, data governance, data warehouse appliances, data warehousing, and master data management.

Here's what you will find inside:

CASE STUDIES

What Works case studies are meant to present snapshots of the most innovative BI and DW implementations in the industry today. The case studies included in this volume demonstrate the power of DI technologies and solutions for industries ranging from prebuilt analytic applications to window coverings to medical research.

LESSONS FROM THE EXPERTS

Included in this issue of *What Works* are articles from leading experts in the services, software, and hardware vendor communities. These lessons provide perspectives about DI best practices and trends.

Q&A WITH THE EXPERTS

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

FEATURE ARTICLE

In this issue, the feature article comes from Philip Russom, senior manager of TDWI Research. In "Data Integration Architecture," he makes a case for the ability of data integration architecture to impose order on the chaos of complexity in data integration.

TDWI RESEARCH

There's more from TDWI Research. *What Works* includes excerpts from TDWI's recent best practices reports: Data Governance Strategies, TDWI's latest report from Philip Russom, and Strategies for Managing Spreadmarts, from Wayne W. Eckerson.

We hope you enjoy this collection of case studies, best practices, and expert insight focused on data integration. 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 that 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

tdwi
www.tdwi.org

Reaching the staff: Staff may be reached via e-mail, telephone, fax, or mail.

E-mail: To e-mail any member of the staff, please use the following form: FirstinitialLastname@1105media.com

Renton office (weekdays, 8:30 a.m.–5:00 p.m. PT)
Telephone 425.277.9126; Fax 425.687.2842
1201 Monster Road SW, Suite 250, Renton, WA 98057

Corporate office (weekdays, 8:30 a.m. – 5:30 p.m. PT)
Telephone 818.734.1520; Fax 818.734.1528
9121 Oakdale Avenue, Suite 101, Chatsworth, CA 91311

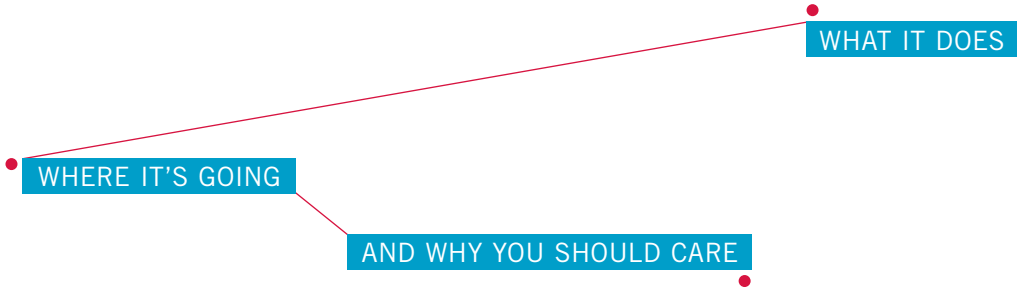
Sponsorship Opportunities: Denelle Hanlon, 425.277.9130, dhanlon@tdwi.org.

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

Table of Contents

Data Integration Architecture



By Philip Russom, Senior Manager, TDWI Research

To a lot of people, the term data integration architecture sounds like an oxymoron. That's because they don't think that data integration has its own architecture. For example, a few data warehouse professionals cling to practices of the 1990s, when data integration was subsumed into the larger data warehouse architecture. Today, many data integration specialists still build one independent interface at a time, a poor practice that's inherently anti-architectural. And a common misconception is that using a vendor product for data integration automatically assures architecture.

Here's the dirty rotten shame of all this: If you don't fully embrace the existence of data integration architecture, you can't address how architecture affects data integration's scalability, staffing, cost, and ability to support real time, master data management, SOA, and interoperability with related integration and quality tools. And all these are worth addressing.

This article makes a case for data integration architecture, to help data integration professionals design and deploy architectures that are strongly independent, future-facing, productive, scalable, and interoperable. The case is made by defining what data integration architecture does, where it's going, and why you should care.

Complexity is the main reason why data integration needs architecture.

This article focuses on the available architectures for relatively complex data integration implementations. In these cases, data integration effects a flow of data from diverse source systems (like operational applications for ERP, CRM, and supply chain, where most enterprise data originates) through multiple transformations of the data to get it ready for loading into diverse target systems (like data warehouses, customer data hubs, and product catalogs). Heterogeneity is the norm for both data sources and targets, since these are various types of applications, database brands, file types, and so on. All these have different data models, so the data must be transformed in the middle of the process, and the transforms, themselves, vary widely. Then there are the interfaces that connect these pieces, which are equally diverse. And the data doesn't flow uninterrupted or in a straight line, so you need data staging areas. Simply put, that's a ton of complex and diverse stuff that you have to organize into a data integration solution.

Goals of Data Integration Architecture

Here's where data integration architecture comes in. It imposes order on the chaos of complexity to achieve certain goals:

Architectural patterns as development standards. Most components of a data integration solution fall into one of three broad categories: servers, interfaces, and data transformations. With that in mind, we can venture a basic definition:

Data integration architecture is simply the pattern made when servers relate through interfaces.

The point of an architectural pattern is to provide a holistic view of both infrastructure and the implementations built atop it, so that people can wrap their heads around these and have a common vision for collaboration. Also, when you inherit someone else's work, you get up to speed faster when they've followed development standards and established patterns. Well-run organizations have development standards, and architectural patterns should be among those.

Simplicity for reuse and consistency. As development standards and architectural patterns are applied to multiple data integration projects, the result is simplicity (at least, compared to ad hoc methods), which fosters the reuse of data integration development artifacts (like jobs, routines, data transforms, interfaces), which in turn increases consistency in the handling of data.

Harmony between common infrastructure and individual solutions. For a solution (like a data flow or a project) to be organized in a preferred architecture, the infrastructure (especially the data integration production server and the interfaces it supports) must enable that architecture.

12 IT systems integrated through 66 interfaces

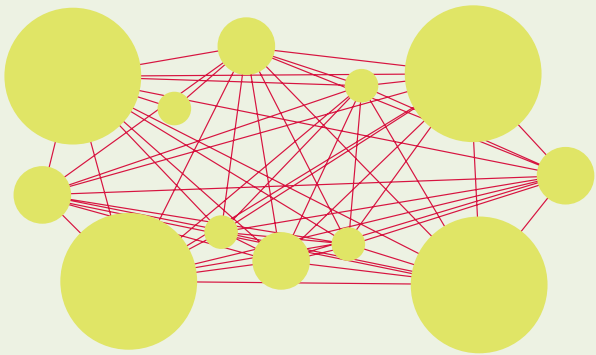


Figure 1. Point-to-point architecture.

12 IT systems integrated through 12 spokes and a hub

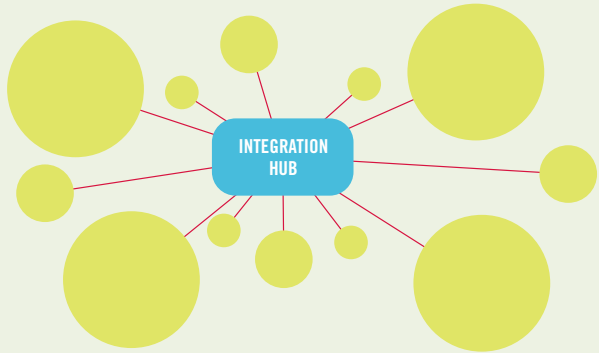


Figure 2. Hub-and-spoke architecture.

Hub-and-spoke is the preferred architecture for most integration solutions.

The most common architectural pattern for data integration is hub-and-spoke architecture. In this architecture, inter-server communication and data transfer pass through a central hub, where an integration server manages communications and performs data transformations. When data integration solutions are built atop a vendor's tool, the server at the hub is usually a vendor's data integration server. With home-grown solutions, the server at the hub may be a database management system or a collection of hand-coded routines. Hybrid systems combine these.

Most integration technologies are today deployed in a hub-and-spoke architecture. This is true of the form of data integration known as extract, transform, and load (ETL). Variations of ETL—like TEL and ELT—may or may not have a recognizable hub. But it's not just ETL. For example, hubs are common in deployments of enterprise information integration (EII). Replication usually entails direct interfaces between databases, without a hub; but high-end replication tools support a control server or other device that acts as a hub. Enterprise application integration (EAI) tools depend on message queue management, and the queue is usually supported by an integration server through which messages are managed.

Benefits of hub-and-spoke architecture

It provides a flexible architectural pattern. The hub-and-spoke concept is easy to understand and work with, yet can be expressed in infinite variations.

It fosters reuse. You typically develop an interface—called a spoke—from the hub to a given system and then reuse that interface as more systems need to communicate with the first one.

It reduces the number of interfaces. The practice of spoke reuse fostered by hub-and-spoke architectures dramatically reduces the number of interfaces you need to build and maintain.

To prove this last point, let's compare hub-and-spoke architecture to its nemesis: point-to-point architecture. This is where IT systems communicate directly without a hub or other remediation. Most interfaces in point-to-point architecture are unique to a specific pair of IT systems and so are not easily reused. Also, point-to-point architecture is often developed through hand coding, which is not productive, thereby raising payroll costs. But the real problem arises when you push point-to-point architecture to an extreme. If you connect every IT system to all others in a collection of n systems, you end up with n!-n interfaces—and that's a lot of interfaces!

For example, Figure 1 illustrates how a collection of 12 IT systems is fully integrated via 66 individual interfaces (12!-12=66). Users often describe the resulting architectural pattern disparagingly as "spaghetti." However, in Figure 2 we see that the same 12 IT systems can be fully integrated through 12 reusable spokes and a hub. The resulting architectural pattern is simple to design and maintain, due to the reduced number of interfaces. This shows how the choice of an integration architecture can impact development standards, reuse, developer productivity and related costs, and the number of interfaces to design and maintain.

Pure architecture is rare; distributed hybrids are common.

The hub-and-spoke concept is a handy symmetrical abstraction, but in the real world only the simplest of integration solutions comply with it 100%.

Point-to-point interfaces can complement hub-and-spoke architecture. Even when integration infrastructure has a hub through which most interfaces communicate, a few point-to-point interfaces that circumvent the hub can be useful. Such warts on the architecture make sense when you just need to copy data from point A to point B, and you don't need the hub's scheduling or data transformation capabilities. Also, a direct interface may be faster than going through the hub. After a bit of evolution, most architectures end up hybrids like this, anyway.

An integration server may become a performance bottleneck. If you keep the hub-and-spoke architecture pure with a data integration implementation, you force all data flows and data processing through a single server. (See Figure 3.) When large data volumes and/or highly complex transformations are involved, it's common to avoid a purely centralized data integration architecture in favor of a distributed one that distributes data processing across more servers to assist with scalability. (See Figure 4.)

Source and target databases may take an active role in a distributed architecture. We think of source and target systems as passively handing data to a data integration implementation and receiving data from it. Yet, source and target systems usually include a database management system (DBMS) that's capable of handling some of the data processing. When a DBMS has available capacity, it makes sense to put it to work pre-processing data before it leaves the source or post-processing data after loading it into a target (which is typical of ELT configurations). Either way, this reduces the load on the data integration server at the hub of the architecture. (See the far-left and far-right sides of Figure 4.)

A distributed data integration architecture may include multiple integration technologies. Besides the role of DBMSs just mentioned, distributed architectures often depend on data quality tools, sort tools, and hand-coded routines to pre- or post-process data, to perform unique operations not suited to the integration hub, or to simply offload the hub.

Scheduling time on the hub is a challenge. For example, a source system may have an optimal window for presenting data that happens to be at a moment when the integration server is engaged. Or you may need to transfer data via flat files. In these cases, the data integration architecture includes one or more data staging areas. This is where data is parked until another process picks it up; data may sit passively or be actively processed before entering the hub or after exiting it.

Data integration architecture is not the same as data warehouse architecture.

The simplified architecture shown in Figure 3 includes an enterprise data warehouse (EDW) as an outbound spoke from a data integration hub. The reality is that an EDW is itself a complex environment that includes many components that need an organizing architecture. For decades, data warehouse professionals have fought a religious war over data warehouse architectures. We all agree that a data warehouse needs an architecture, although we can't agree which one!

Without taking a side in the EDW architecture debate, let's note that most recent definitions of data warehouse architecture focus on data models and how they're expressed in appropriate types of databases. Sometimes the resulting architecture has a hub, sometimes not. That's different from data integration's architectural focus on interfaces and data transformations, which almost always hinge on a hub. Obviously, data integration architecture must feed data

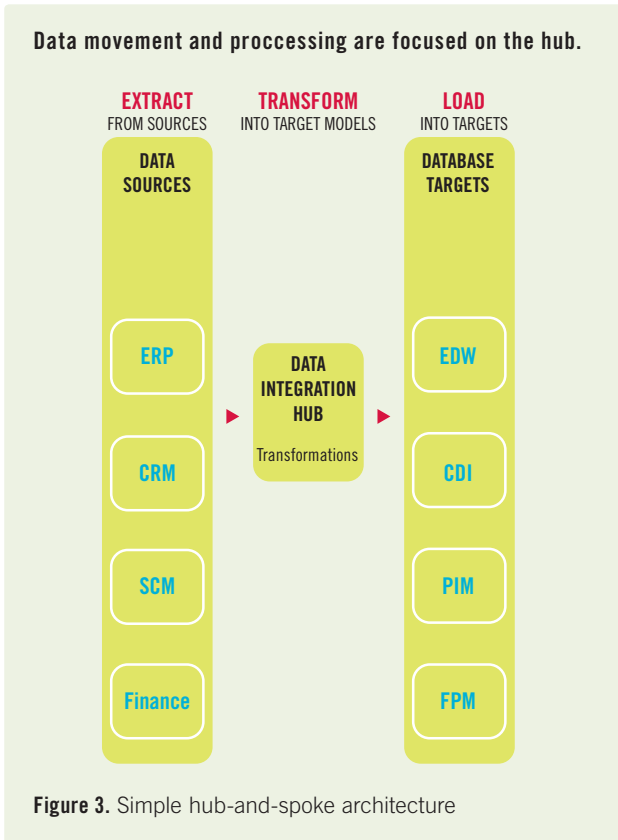


Figure 3. Simple hub-and-spoke architecture

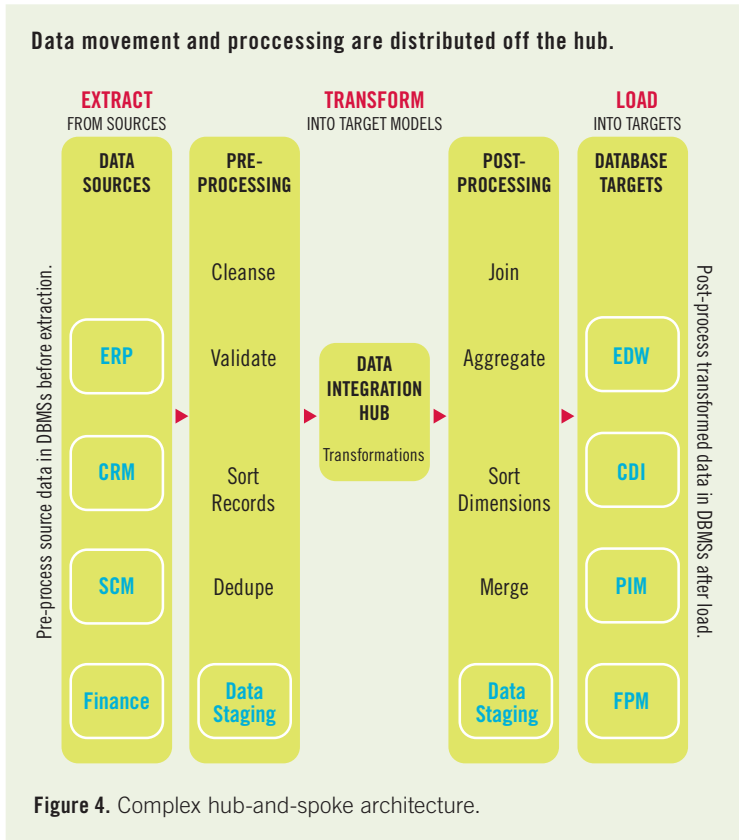


Figure 4. Complex hub-and-spoke architecture.

into data warehouse architecture, so they overlap. Yet, the two have different foci and patterns and so should be considered separate.

Data integration is an autonomous practice, which needs its own architecture.

Many of the best practices and technologies of data integration originated in data warehousing in the early 1990s, and data integration continues to be a meaningful and growing segment within data warehousing. For these reasons, some data warehouse professionals continue to think of data integration architecture as a subset of data warehouse architecture. Yet, by the turn of the current century, data integration had begun its journey into independence and today should be considered an autonomous practice. Two recent trends corroborate this independence.

Operational data integration. This involves the migration, consolidation, synchronization, and upgrade of operational databases and applications. In other words, it's a data integration practice that doesn't involve data warehousing or business intelligence, the way that analytic data integration does. According to TDWI Research, both operational data integration and analytic data integration are growing, yet the operational practice is growing faster than the analytic one.

Data integration's life outside data warehousing has corroborated its independence in recent years but also forced changes to how it's funded and staffed. A struggle many organizations face is the fact that data integration specialists are usually members of a data warehouse team, which has specific funding and staffing. Pulling them away from analytic data integration work so they can do operational data integration work creates organizational conflicts.

Data integration competency centers. To avoid conflict, to assure that all work gets done, and to avoid redundant teams and infrastructure, many organizations have founded a data integration competency center. A competency center is a neutral organization that provides centralized shared services in support of a range of business initiatives and technical implementations—not just data warehousing. Furthermore, relative to architecture, a data integration competency center establishes development standards and architectural patterns; it encourages reuse; and it owns enterprise data integration infrastructure (which influences architecture). The rise of an independent team—as seen in competency centers—is further proof of data integration's autonomy.

Data integration architecture is set to go service-oriented.

Data integration architecture is heading out on the leading edge by incorporating service-oriented architecture (SOA). Note that SOA won't replace current hub-based architectures for data integration. Hubs will remain but be extended by services. The goal is to provide the ultimate spoke, namely the data integration service. For that to happen, the integration server at the hub has to support a data integration service registry.

The benefits of SOA to data integration architecture are enormous. After all, data integration is all about interfaces to heterogeneous systems coupled with sophisticated data transformation at the hub. SOA gives data integration a wider range of interface possibilities, and many of these allow it to participate in composite application architectures. For example, a data integration service can bring integrated data into a wide variety of applications, especially those for operational BI, embedded reporting, performance management, and dashboards. Depending on several factors, a data integration service may integrate data on demand for time-sensitive business practices like just-in-time inventory or customer service.

A data integration service is a generalized interface, so a data integration tool can call and be called in a reusable fashion from data quality or application integration tools, achieving greater interoperability with these. Progressively, data integration infrastructure is an enabler for data management practices like master data management (MDM), customer data integration (CDI), and product information management (PIM). A data integration service could provide functions for these that are easily embedded in various operational or analytic applications. As you can see, the architectural variations are increasing—in a positive way—as data integration embraces services.

Recommendations

Recognize that data integration architecture exists. Although it overlaps with data warehousing architecture and interacts with the entire business intelligence technology stack, data integration architecture is an autonomous structure demanded of an autonomous practice.

Give the autonomous practice of data integration autonomous staffing. Instead of scavenging the data integration specialists from your data warehouse team, consider establishing a data integration competency center.

Adopt hub-and-spoke architecture for most data integration implementations. After all, the hub reduces the number of interfaces and provides a pattern that everyone can understand and be productive with. And hub-and-spoke architecture is conducive with other worthy goals, like reuse, productivity, collaboration, and consistent development standards.

Don't be doctrinaire about hub-and-spoke architecture. Otherwise, you'll heap a heavy workload on the hub. To accommodate large data volumes and/or complex transformational processing, distribute the workload beyond the hub through various types of pre-processing and post-processing.

Embrace services. The data integration service extends existing hub-and-spoke architectures with new interfaces, so data integration hubs can embed functions into a wide range of traditional and composite application architectures. ●

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, data governance, data warehouse appliances, data warehousing, and master data management (MDM). What do these terms mean, and how do they apply to your organization?

DATA INTEGRATION

page 7

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.

DATA GOVERNANCE

page 28

Data governance is usually manifested as an executive-level data governance board, committee, or other organizational structure that creates and enforces policies and procedures for the business use and technical management of data across the entire organization. Common goals of data governance are to improve data's quality; remediate its inconsistencies; share it broadly; leverage its aggregate for competitive advantage; manage change relative to data usage; and comply with internal and external regulations and standards for data usage. In a nutshell, data governance is an organizational structure that oversees the broad use and usability of data as an enterprise asset.

DATA WAREHOUSE APPLIANCES

page 30

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.

DATA WAREHOUSING

page 34

At the highest level, designing a data warehouse involves creating, manipulating, and mapping models. These models are conceptual, logical, and physical (data) representations of the business and end-user information needs. Some models already exist in source systems and must be reverse engineered. Other models, such as those defining the data warehouse, are created from scratch. Creating a data warehouse requires designers to map data between source and target models, capturing the details of the transformation in a meta data repository. Tools that support these various modeling, mapping, and documentation activities are known as data warehouse design tools.

MASTER DATA MANAGEMENT

page 38

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.

LESSON FROM THE EXPERTS

Building an Agile and Trusted Data Foundation

By Philip On
Director, Enterprise Information Management, Business Objects, an SAP company

As the volume and sources of your organization's data grow exponentially, the need for trusted and accurate information accelerates. To effectively gain insight from your enterprise data, you need an enterprise information management (EIM) strategy that simultaneously addresses data integration and data quality.

Problem

Most organizations approach data integration and data quality projects tactically. A team of developers or consultants assembles to address a specific business need, and when the project is over, they disband—with the result that skills and knowledge are not reused. Compounding the inefficiency are enterprise-class data integration tools that are overly complex and that require months of ramp-up time to be proficient. On the flip side are tools that are too limited—tools that, for example, support advanced requirements such as data cleansing with hand coding.

Solution

Think Strategic

Building an agile and trusted data foundation requires a strategic approach to information management. Think big—but start small. Although you can approach your information management strategy with one building block at a time, you must consider the big picture every step of the way. For example, you may want to start with a data quality project to clean up your customer data during the extraction, transformation, and load (ETL) process. Subsequently, you may want to leverage the same cleansing rules to improve your customer data at the operational customer relationship management (CRM) source system.

Create Standards

Consider forming a competency center for data to drive your organization's data governance and standards. A bank in Belgium faces challenges with getting an accurate view of its business because it uses 15 different terms to describe company revenue, including "total sales," "net revenue," and "net income." A data competency center reconciles the differences and enforces standard terms that ensure more accurate use of information among users. A data competency center helps you resolve data quality and integration issues and, as a result, you deliver trusted data across the organization.

A data competency center helps you resolve data quality and integration issues and, as a result, you deliver trusted data across the organization.

Seek Out the Right Tools

Organizations understandably favor automated tools over hand coding to support data integration and quality projects. These are a few key considerations when selecting a tool.

Prioritize Ease of Use

If effective use of a tool requires hiring specialized consultants, then you should think twice before considering it. A tool should accelerate development cycles, not extend them. Some technologies require separate interfaces for project management, data profiling, ETL design, data quality, debugging, and deployment. Look for a tool that will help you accelerate time to market by allowing you to seamlessly develop an entire data integration or data quality process using a single interface.

Prioritize Data Quality

Data quality is the hardest but most important part of information management. When evaluating technologies, make sure your choice handles rich cleansing processes as well as data in any industry or operational domain, such as customer, product, and beyond. Your data quality solution should allow you to parse, cleanse, complete, match, and consolidate enterprise data anywhere in your IT infrastructure. You should be able to cleanse data at its source, during the ETL process, or during data entry.

Prioritize Metadata Management

Metadata provides the contextual information to help you understand what happens to your data as it passes through various stages of its life cycle—from creation, to transformation, to consumption by the business user. Without metadata, it's impossible to fully answer the question, "Where did this number come from?" To achieve this kind of insight, you need to ensure that the tool you select provides out-of-the-box metadata integration among your data integration, data quality, business intelligence (BI), and enterprise resource planning (ERP) environments.

Conclusion

Building an agile and trusted data foundation begins with a strategy for information management that promotes standards and reuse of resources across the enterprise. Be sure you evaluate your tools on the merits of ease of use, data quality, and end-to-end metadata management. ●

For a free white paper on this topic, download "The Importance of a Single Platform for Data Integration and Quality Management" or "A Roadmap to Data Migration Success," or click here for more information about Business Objects, an SAP company.

CASE STUDY

Hunter Douglas Streamlines Complex Supply Chain

Commentary by Aart Van Leeuwen
Manager, Information Systems Service, Hunter Douglas Europe

Hunter Douglas is the world market leader in window coverings and a major manufacturer of architectural products. Its head office is in Rotterdam, The Netherlands. Hunter Douglas operates its own trading exchange, where it buys and sells aluminum alloys and bauxite. Its global operations range from sites that work with raw materials to distribution centers of finished window fashions. The Hunter Douglas Group comprises 166 companies with 65 manufacturing and 101 assembly operations and marketing organizations in more than 100 countries; it employs about 20,000 people and had sales in 2006 of USD 2.6 billion.

Hunter Douglas manufactures and assembles against customer demand, making its supply chain extremely complex and far-reaching. Traditionally, it has conducted business with its supply chain by phone, fax, and e-mail—inefficient communication methods that created long delays due to numerous time zones and vast amounts of information to share accurately.

> "We could not tolerate anything that would interrupt or cause a slowdown in our production environment, which depends totally upon SAP maintaining its high levels of performance. Fortunately, we discovered Data Integrator from Business Objects, an SAP company. It was clear that Data Integrator offered us the perfect link between the Web and the SAP system. It also proved that it could provide us with the speed we desired—without impacting the performance of the ERP system."
>
> **Aart Van Leeuwen**
> Manager, Hunter Douglas Europe

Challenge

A Need for Greater Integration throughout the Supply Chain

Hunter Douglas deployed SAP R/3 to run the main production process: planning and managing order intake, purchasing, supply, and, of course, financials. To deliver a complete business solution, Hunter Douglas needs to provide rapid and easy access to critical information located in its enterprisewide SAP R/3 and other back-office systems.

Additionally, SAP is critical to Hunter Douglas's strategy for greater integration throughout its supply chain. The company's European headquarters is currently the only site outside of the U.S. that relies upon SAP; other distributors and manufacturing and assembly sites use a variety of ERP and back-office systems.

Aart Van Leeuwen, manager of information systems service at Hunter Douglas Europe is responsible for the European headquarters' e-business strategy and implementation. "Our supply chain is immense and highly complex in its structure and interaction. Our e-business strategy initially needed to benefit both Hunter Douglas's internal and external customers on a global scale," stated Van Leeuwen. Order placement, confirmation, and supply between manufacturing and assembly plants were still manual processes. As orders arrived, they were re-keyed into the order processing system.

"We were looking to cut administration time from hours to minutes," said Van Leeuwen. "We could do this only by integrating requests coming across the Web with our live ERP environment." This is where Hunter Douglas perceived a rapid return on investment. The complexity of its heterogeneous global IT infrastructure also needed to be integrated. "Each company in our group takes an independent and entrepreneurial approach to their IT infrastructure," commented Van Leeuwen.

Approach

Data Integrator: The Perfect Link between the Web and SAP R/3

Data Integrator, offering one of the few ways to rapidly and intelligently catch information, provides rapid deployment for a fast return on investment. It enables the extraction of information from any combination of ERP, supply chain management, or other enterprise application. Data from these applications is critical to an enterprise as it

generates constantly changing data, affecting inventory levels, account details, delivery dates, and financial reporting. Assembly sites were provided with a direct-connect Web site where staff can build and validate custom configurations, check prices, receive accurate promise dates for delivery, and check the precise status of outstanding orders in real time. With this stage of the supply chain automated, assembly plants gain faster access to order information.

Results

Improvements in Service Speed and Efficiency

The deployment of Data Integrator at Hunter Douglas has revolutionized how the company accepts, processes, and confirms orders. "Our customers can now submit orders straight into our system and receive order confirmation," Van Leeuwen said. "Behind the scenes, Data Integrator obtains the data, translates it into XML, validates it, and confirms delivery timelines."

"We have been marketing this capability to our larger customers, those that place bulk orders," Van Leeuwen added, "and this has been a resounding success, since they have realized improvements in both service speed and efficiency." ●

HunterDouglas®

Hunter Douglas manufactures and assembles to customer demand, making its supply chain extremely complex. Data Integrator provides an ideal solution to automating the order and fulfillment processes.

BUSINESS PLAN

Advance and streamline its highly complex supply chain to:

- Automate order, planning, and fulfillment processes
- Improve the efficiency of worldwide assembly operations
- Reduce timeframe for confirmations from hours to minutes
- Improve customer satisfaction

WHY BUSINESS OBJECTS, AN SAP COMPANY?

- Manage and automate more than 10,000 orders a day of highly configurable products from 300 distributors on a worldwide basis
- Instant promise to deliver dates, order and delivery status, and invoice reconciliation
- Improved response time—from more than one minute to less than three seconds
- Increased level of customer satisfaction and employee productivity

CASE STUDY

UMIT Fights Cancer with Data Integration, Mining, and Analysis

Commentary by Dr. Bernhard Pfeifer
Associate Professor, University for Health Sciences, Medical Informatics and Technology

The IMGuS project

Prostate cancer is the most frequent tumor type in males and the second most frequent cause of male death. The IMGuS (Institute for Medical Genomics Research and Systems Biology) project aims at the application of high throughput data processing to identify molecular signatures allowing the stratification of patients who are susceptible to curative treatment of prostate cancer and who need treatment.

A key participant in the IMGuS project, the University for Health Sciences, Medical Informatics and Technology (UMIT), based in Hall (Austria), manages the technical infrastructure and the life science data warehouse part of the project, in coordination with five other research groups.

Data processing is key to cancer research

"A large part of cancer research today consists of data processing and statistical analysis," explains Dr. Bernhard Tilg, professor and board member at UMIT Institute of Biomedical Engineering. "The goal of these projects is to identify molecular signatures associated with certain types of tumors, so that efficient and non-intrusive diagnostic mechanisms can be designed. Some cancer treatments have high success rates, when the disease is diagnosed in time, but the key problem remains the diagnostic."

"We use data integration to combine several different data sources to perform advanced analysis and statistics on the whole set," clarifies Dr. Bernhard Pfeifer, associate professor at UMIT Institute of Biomedical Engineering. "And because of the amount of data the high throughput sources create, an automated approach is mandatory. We looked at a number of data integration solutions, both proprietary and open source, and settled on Talend's solutions because of their flexibility, openness, and high performance."

> "UMIT/biomed relies entirely on Talend's solutions for all data integration needs. We have high hopes that the IMGuS project will contribute to the reduction of prostate cancer mortality rates, and data integration is a critical part of this project. Talend is helping us save lives!"

Indeed, it is critical for the project that the chosen data integration solution not only work with all data sources, but also be able to integrate specific data processing approaches; for example, since various medical devices deliver data in different formats, preprocessing of this data is required. Talend's open architecture allowed UMIT to develop specific components to access and process this data.

The PostgreSQL-based LINDA data warehouse, which is the basis for the statistical analysis of the IMGuS project data, is loaded in two stages. The first stage, dubbed electronic data capture, or EDC, centralizes data from all the different sources: patient samples, reference medical data, genome cartography, etc. "The complexity of the electronic data capture stage is very high," explains Pfeifer. "Not only are the data providers very diverse—five different universities and research centers—but the formats vary vastly: very large CSV files, high resolution images, RDBMS, XML data, etc."

Administrative data is also loaded at this stage: patient demographics, information about the biological source a certain sample comes from (tissue, serum, etc.), or information on the data source in which the information is stored.

The second loading stage reconciles, transforms, cleanses, and enriches the data contained in the EDC and loads the LINDA data warehouse. "At this stage, we need to bring in reference data from external providers—medical publications, legacy systems, reference medical databases. Talend's native support of Web services and XML brings tremendous value to the project," Pfeifer says. "It allows very easily to parse and to cross reference external sources of data, reducing greatly the time it would otherwise take to enrich the data warehouse."

The frequent refresh of the data warehouse, performed every night, ensures that researchers can use ad hoc query and data mining tools and apply advanced statistical models to extract data relevant for their research.

"UMIT/biomed relies entirely on Talend's solutions for all data integration needs," concludes Pfeifer. "We have high hopes that the IMGuS project will contribute to the reduction of prostate cancer mortality rates, and data integration is a critical part of this project. Talend is helping us save lives!" ●

For a free white paper on this topic, download "Integrating Data in the Information System: An Open Source Approach," or click here for more information about Talend.

LESSON FROM THE EXPERTS

Open Source: Beyond the IT Infrastructure, Into the Business Applications

By Yves de Montcheuil
Vice President of Marketing, Talend

Open source on every floor

Open source, traditionally viewed as developer and infrastructure tools, is now infiltrating the enterprise IT environment.

We already know that open source software is widely present within companies' infrastructures: security (firewall, IPS-IDS, sniffer, proxy, antivirus, anti-spam, etc.), operating systems (workstations, network, scientific computers, etc.), databases, Web browsers, etc.

Today, open source technology is found in the lower layers of almost all information systems. It is also deployed in the higher layers (business applications) as well as the middleware layers (not visible to the user), like Talend Open Studio, Talend's flagship data integration product.

Offerings are maturing

The growing presence of open source software in the information system illustrates the greater maturity of the offerings. Some years ago, open source vendors were "preaching to the choir" by offering their solutions to users that already supported the open software movement. The contrast today, however, is that adopting organizations select open source solutions because of the competitive advantage they deliver.

Every day, new companies—including some of the largest in the world—announce their decision to develop a solution under an open source license. For government organizations that generally have stricter budgets and resources, cost and administration considerations have traditionally played a large role in IT decision making. Private businesses are not far behind. The International Oracle Users Group recently conducted a survey, which shows that 37 percent of companies using an Oracle database are also using an open source database.

The benefits of open source

With the success and growing adoption of open source clearly established, we turn our focus to its unmistakable advantages:

Quality and reliability. A long-standing pillar of the open source community is the emphasis on reliability of the code and the ability of the community to constantly fine tune the applications. While proprietary vendors seek ongoing customer loyalty, open source providers seek customer satisfaction.

Transparency. Open source code provides open access to anyone. Not only does this make the customization easier but it also reinforces the users' independency toward the proprietary world.

Cost savings. Numerous debates have highlighted the differences between open source and free software. Although the cost savings have attracted most companies to use these technologies, it is not the main argument in favor of open source.

Respect of standards and interoperability. For years, interoperability has been one of the main challenges that faced software users. Proprietary solutions, which have been developed by isolated teams to answer specific needs, have not been designed to share, converse, and collaborate. Vendors need to alter their line of thinking and shift from locking in users with proprietary specificities and instead open up their solutions to facilitate integration into information systems.

talend*
*open data solutions

Another advantage often cited is the **reduction of the strain on resources**. Indeed, open source solutions do not generally need high performance systems to deliver great performance, but they can contribute to leveraging existing systems. This is in addition to the already analyzed economical advantages.

And the drawbacks? The main drawbacks that are commonly linked with open source solutions are slowly disappearing. The main argument against open source in the past has been the sustainability of open source. Today, however, the natural evolution of successful open solutions and the technical support benefits from well-established and reliable structures have changed these perceptions. Besides, open source vendors are receiving increasing support from private investors and venture capital.

The future

Open source is becoming solidly established in enterprise IT—with even more widespread adoption on the horizon. In less than five years, many open source vendors have moved from marginal positions to being some of the most established players alongside prestigious companies and institutions. ●

CASE STUDY

Operational ETL Provides Solution for Past Data Quality Woes

By Chris Jennings
Senior Principal, Collaborative Consulting

An educational publishing company (EPC) wanted to replace its core systems, including finance, order management, e-commerce, and content management. Early in the initiative, it was determined that a key element for success would be a data quality program instituted as part of the data conversion and integration effort. Experience and research into the common causes of failure for large ERP implementations highlighted this as a substantial risk area.

A consistent theme that was echoed in the failed implementations was the inability to properly cleanse, consolidate, and restructure data as part of the data conversion and ongoing data integration processes. This finding was not surprising, as it is the point of failure for many business intelligence initiatives. These difficulties commonly result from one or both of the following problems: improper and complicated technology solutions, or a lack of focus on understanding the data properly. This case study focuses on the technology solution.

EPC chose to standardize on an ETL tool (Informatica) as the backbone of the data integration architecture. While many practitioners relate ETL tools with batch data warehouse architectures, ETL tools increasingly are used for operational systems integration efforts. The base transformation functionality in the leading ETL tools has become mature; therefore, major changes in transformation functionality from release to release are not common. This has allowed vendors to grow in other areas, such as SOA enablement and stronger integration with data quality technologies.

These changes have made the ETL tools more attractive for broad use in handling the majority of data integration tasks across the enterprise. EPC is utilizing the base Informatica ETL tool PowerCenter with the real-time and Web services options, Informatica Data Quality, and PowerExchange for Oracle E-Business Suite.

The ETL backbone envisioned by EPC for the first release of the ERP project satisfies the required architecture functions. It was quickly evident that the functions in an operational environment are not significantly different from those in a data warehouse environment. Here are some examples of the functions:

- Flag, track, and fix errors
- Cleanse name and address data
- Schedule jobs
- Read and write from heterogeneous sources and targets
- Translate source systems codes into standardized codes
- Handle change data capture
- Transformation to restructure data

While the necessary functions are the same, the implementations are different from those in a data warehouse. Here are a few examples of these differences:

Error records in a data warehouse are flagged and tracked but are seldom handled or fixed because they do not have significant impact. Based on aggregate analysis, a single record missing is often statistically inconsequential. In an operational system, this is not true. For example, if an incoming order record fails because a product cannot be found, there is a customer waiting for the product to be shipped. The error must be corrected in a timely fashion.

These changes have made the ETL tools more attractive for broad use in handling the majority of data integration tasks across the enterprise.

Addresses in a data warehouse often are cleansed in batch. At EPC, the operational system needs them cleansed in real time. The SOA enablement of the ETL technology has allowed EPC to utilize the ETL backbone for this function.

In data warehousing environments, the ETL tool or enterprise scheduling packages are used to execute large batches with many jobs having extensive dependencies. These batches are often run overnight. For EPC, we are using the ETL tool to integrate data from multiple applications using smaller batches that run frequently throughout the day. ●

For a free white paper on this topic, download "Master Data Management: Creating an Enabling Platform for Business Integration," or click here for more information about Collaborative Consulting.

LESSON FROM THE EXPERTS

Data Integration: Consider It an Enabling Platform

By John Williams
Senior Vice President, Collaborative Consulting

Introduction

Historically, data integration was often considered an activity to pass information between applications. More recently, data warehousing highlighted the need for more substantial data integration capabilities, as it required integration across multiple applications and consolidation into a single environment. While the two types of integration may differ in some ways, they are fundamentally alike, with the same basic requirement: get key elements of information moved from one application to the other and transformed into a format that the receiving application can use.

Data integration also has presented one of the key challenges to any new application implementation. As much as one half of the effort expended on new application implementation is focused on redeveloping the integration points to other applications. Much of this effort is spent on understanding the specific details of the existing integrations before beginning the rewrite process. This occurs regardless of whether the integration is to a data warehouse or an operational application.

The Data Integration Platform

This traditional model is changing rapidly. The current focus on information as an asset, as well as the rapid evolution of data integration tools, has changed the way we look at data integration. As organizations move to new application platforms, implement new data warehouses, or simply struggle with new application implementation, they are beginning to realize that having a robust data integration capability is key to their continued success.

More and more organizations are envisioning and creating a data integration platform. This concept has been evolving over the past decade or so with the advent of data warehousing and ETL tools, which fundamentally have been focused on batch processing. It has been fostered further by the introduction of EAI and messaging architectures, which are more focused on real-time and near-real-time data movement. It is not simply a technology; it also consists of process, people, and tools.

The current focus on information as an asset, as well as the rapid evolution of data integration tools, has changed the way we look at data integration.

Creating a data integration platform enables an organization to provide a faster and more flexible mechanism for data integration. This environment can be used across both operational and analytical applications, and its sole purpose is to facilitate information movement between environments. It limits or eliminates redundant efforts to map data from one system to another by creating a documented understanding of the information that is being provided by one system to another, enabling other systems that need that same information to take advantage of its availability. It provides a methodology for making consistent decisions on how to integrate data across existing and new applications and creates institutional knowledge of data integrations so that they can be maintained over time.

Application Platform Vendor Offerings

This capability is fast becoming one of the key selling features of the application platforms. All of the major solution providers have made rapid application integration a crucial selling feature of their particular solution, most recently adding business intelligence integration to their portfolios. The current risk with the platform provider offerings is that they do not consider integration to products outside of their offerings adequately. Unless an organization is vested fully in a single vendor offering, it will be critical for it to create a data integration platform independent of the platform vendor offering to allow for integration of other applications.

Conclusion

Organizations must quickly develop a data integration competency or continue to struggle and spend valuable resources reinventing application integrations every time a new application comes on board or an existing application needs additional data. The tools have evolved and become substantial enablers to creating this capability; but beware of assuming that the application platform vendor offerings are adequate to the task. Focus on creating a data integration platform, made up of people, process, and tools, that provides a capability that is an advantage to the entire enterprise. If done right, this capability will not only accelerate application implementation initiatives but also foster an institutional knowledge of information that can be leveraged across the enterprise now and in the future. ●

CASE STUDY

Keeping Information Fresh at Utz Quality Foods

Commentary by J. Ed Smith
Chief Information Director, Utz Quality Foods, Inc.

Utz is a national player in the consumer packaged goods industry. The company services large retail customers, such as Sam's Club and Costco, as well as major supermarket chains. It produces and delivers more than 20,000 pounds of potato chips every hour and manages more than 700 delivery routes to 30,000 stores.

To coordinate this massive operation, Utz Quality Foods uses portable business intelligence (BI) and data integration technology to streamline sales, marketing, production, and distribution activities.

A New Recipe for Success

The company's route salespeople collect data using handheld devices. They have about 50 transactions available in a portable information system, which includes buying back product that is no longer fresh and selling new product into the store.

Pertinent information is recorded in the handhelds as the route salespeople make their rounds. Upon returning to their local distribution centers, the salespeople simply dock the portable handheld computers; then, through various telecommunication methods, the data is uploaded to a DB2 database on an IBM iSeries 520 computer.

The database is updated daily with current inventory and demand information from all of the salespeople. Each night, the company processes the information in batches. A separate application sorts and massages the data into meaningful tables that are structured for efficient hierarchical reporting. Fresh information is available each morning for reporting and analysis via a BI environment called UtzFOCUS, built with Information Builders' WebFOCUS BI software.

Spicing Up Sales

Using UtzFOCUS, managers can quickly gather and analyze current sales and distribution information. For example, a regional manager can see if a store has not been properly serviced, then find the source of the problem, and correct it quickly. The system also delivers daily sales breakdowns—product-by-product and store-by-store—in a form that is easy for busy salespeople to digest. Users can determine how much of a certain product is selling on any given day, how much was sold into a particular store or chain, at what price, and in response to which strategic promotion.

Putting real-time information into the hands of the people who need it most is an important part of the operation. Decision makers at the company look at sales information every day to get a current view of demand, not just at month's end to analyze sales. Regional and district managers use the reports to make daily decisions about inventory and sales. For example, often they compare the previous week's sales results to the sales from the same week a year earlier, drilling down to isolate areas of interest. UtzFOCUS enables the company to root out problems such as lagging sales in a particular region or store. All of the information is accessible via standard Web browsers.

A Taste of Things to Come

The company has proven that it can synchronize supply and demand by quickly exchanging information between headquarters and the field. Currently, it's exploring ways to tie daily production more closely to sales in order to minimize reserves of bags, shipping containers, and other supplies—and thus allow the factories to maintain just enough inventory for current needs.

Utz Quality Foods is also discovering ways to deliver portable analytics to the sales team using a technology called WebFOCUS Active Reports, which will allow salespeople to manipulate the DB2 data offline. Instead of simply receiving standard reports from the daily sales data, users will be able to interact with the data to make decisions on the spot.

Thanks to this system, the company is growing faster than any of its competitors in its core markets and products. The BI environment and associated data integration tools continue to increase the effectiveness of sales promotions and to improve customer relationships. ●

For a free white paper on this topic, download "Worst Practices in Business Intelligence: Why BI Applications Succeed Where BI Tools Fail," or click here for more information about Information Builders.

LESSON FROM THE EXPERTS

Data Warehouse Alternatives: Seven Data Integration Options for BI Solutions

By Kevin R. Quinn
Vice President of Product Marketing, Information Builders, Inc.

While data warehouses are important for many types of analytical systems, many BI applications are better served with data integration technologies that pull data into reports as needed. There are seven basic ways to integrate and access data to solve various business problems.

1. Traditional Data Warehouse

Data warehouses traditionally involve gathering data from multiple sources to create an aggregated source of information for reporting. Information is extracted from production data sources as it is generated (real-time information), or in periodic stages (latent information). It is often simpler and more efficient to run queries against this data, rather than to access each data source separately. Traditional data warehouses work well when you need to reduce overhead on a transaction-processing system, analyze historical data that is no longer accessible in operational applications, or aggregate data from multiple sources.

2. Real-Time Data Warehouse

Real-time data warehouses are constantly updated by "trickle-feeding" data from production data sources, rather than uploading data in batch mode at periodic intervals. This is a good approach when you need current data in your reports. Instead of migrating data from operational systems into a central data warehouse, you can use real-time integration technology to deliver the data whenever it is entered into operational systems.

3. Operational Data Access

Operational business intelligence systems give users a real-time view of business events as they occur, such as shipping orders to customers, routing parts through an assembly line, or sending trouble tickets to customer service reps. These applications generally obtain information from an automated workflow process or directly from production systems. There is less latency between when an event occurs and when the BI system is aware of that event, putting business users in touch with current information.

4. Enterprise Information Integration (EII)

Enterprise information integration (EII) refers to the real-time aggregation of data across multiple data sources. EII solutions present distributed data as if it exists in a single location. This distinguishes EII from other types of data access technologies, since data is not permanently moved or replicated into a new location or database. The source data remains intact.

5. Process Integration

Users querying a database or running a report typically initiate analytical BI systems. However, a BI system can also be triggered by a business process. For example, when an ERP system receives an order or a manufacturing process updates a bill of materials, these events might notify other applications. In some cases, users are asked to supply input. In other cases, there is no user input involved.

6. Search Technology

Most search engines are designed to index and track Web pages, not database transactions. However, with the right integration technology, you can use search technology to unleash information that is locked up in proprietary information systems as well. This enables users to search dynamic business intelligence content in addition to structured and unstructured data sources, and to create Google-style results from data sources throughout the enterprise.

7. Web Services

With the right Web services adapter, you can treat data coming from an Internet Web service as if it were stored in a relational table. This offers many reporting and analytical options—without recourse to a data warehouse. For example, a purchasing officer might need to review a supplier's inventory, pricing, and delivery options to determine which items to restock. If that information is available as a Web service, the officer could retrieve it in a single report and make an instant re-stocking decision.

Summary

A complete integration platform can streamline all of these data integration projects by providing a cohesive solution for extract, transform, and load (ETL) procedures, EII initiatives, Web services deployments, and many types of business process integration scenarios. Analyze each business challenge to understand whether a data warehouse or another type of information-access method presents the best solution. Always try to identify the best method at the outset of the project, and don't assume that a data warehouse is the correct solution before assessing all the options. ●

CASE STUDY

A Global Financial Services Company Signs Up for Speed

By Lia Szep
Senior Technical Analyst, Syncsort Incorporated

We've all gotten the e-mails: a three-night stay on some island that you've never heard of, $50 off the latest everything-but-the-kitchen-sink cell phone, or a six-day cruise to anywhere. If you happen to love Bora Bora this time of year, want a cell phone that walks your dog, or think "anywhere" is better than here—you may be thinking "sign me up." Believe it or not, there is big money to be made in people who want to go to Bora Bora—if you can find them. One century-old, global financial services company (which we call GFS) is doing just that.

Background

A data management group within GFS has been tasked, not with getting new customers, but with keeping current customers engaged. The group works to identify models of customer behavior by applying variables to customer data sets. They then run campaigns designed to cross-sell or up-sell with touch points that include e-mail, Web, and statement inserts. A refresh process uses the results of these campaigns to produce new variables, which are used to identify new models, which are the basis for new campaigns. And so runs the information delivery continuum.

Challenge

For a company like GFS, with 30-40 million active customers, 10,000 variables each, and double-digit terabytes of data, crunching these numbers was no small task. Using SAS to process everything—from the variables to the models to the data sets to the campaign lists—was taking 30 days. The refresh process, using DB2, which accounts for 95% of the feeds, was averaging 14 hours. They also have stringent SLAs. "If we miss the Monday deadline, we lose millions," says the technical architect at GFS.

GFS Cuts Campaign Time by 83% with DMExpress

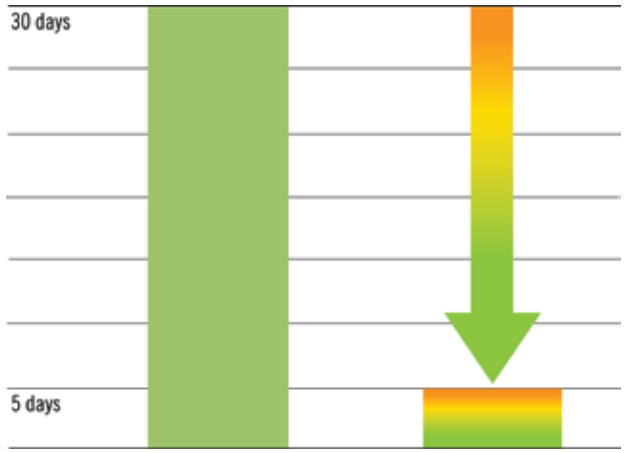


Figure 1. A campaign that once took 30 days to complete now takes five.

Solution

Facilitated by Syncsort's DMExpress, the group went to work developing a confidential new architecture. Once in place, DMExpress was used to build the feeds and replace much of the SAS processing that created the data sets and campaign lists. A campaign that once took 30 days to complete now takes five. And with DMExpress now building the feeds, the refresh process is also significantly faster. The process that once took 14 hours now takes two. "We are in the position, using DMExpress and the new architecture, to run customer behavioral models in a half a minute," says the architect. "Campaigns we were only able to run once a month, we now run four times a month. And that is only because our SLAs are once a week. If business users wanted it, we are positioned to run even more campaigns."

In addition, GFS has benefited from DMExpress's ease of use. For a group of command line users, the easy-to-use graphical user interface means more flexibility with staff. "With DMExpress, we can get new people up to date in two to three days," the GFS architect says. "Now, when we have turnover, we don't have to lose a base of valuable knowledge."

DMExpress is currently being considered for other projects at GFS. And with the success easily measured in greater customer profitability, the architect, who runs "one of thousands of data warehouses" in the company, remains a staunch supporter of Syncsort and DMExpress. ●

For a free white paper on this topic, download "Optimizing ETL: The Key to Accelerating Data-Driven Business Decisions," or click here for more information about Syncsort Incorporated.

LESSON FROM THE EXPERTS

The Wheel Keeps Turning

By Lia Szep
Senior Technical Analyst, Syncsort Incorporated

The more successful an enterprise becomes, the more data it generates. The more data it generates, the harder it is to reap actionable data. Actionable data sustains business intelligence. Business intelligence supports enterprise success. The more successful an enterprise becomes, the more data . . . feel like a hamster? Consider this: If you were to wipe away the bells and whistles and sort through the maze of processes and systems that have been developed to answer both large and small corporate questions, what would you be left with?

Information delivery comes down to two things: data integration and business intelligence. And with two major players on the field—data producers and data consumers—a company's success comes in part from its ability to get data from the hands of one to those of the other. The faster you can process customer and market data, the better you can anticipate and respond to changing business trends. The key to getting the most out of your BI solution is finding the right DI tool. This isn't all that difficult, if you keep in mind a few basic objectives.

Cut Processing Time

The purpose of a BI solution is to enable business users to make crucial business decisions. Unless you can minimize the time it takes to complete a job, you have no chance of meeting this objective. The faster a DI tool, the more timely and actionable the data; so look for a solution that speeds querying and quickly creates and loads aggregate tables.

Reduce Total Cost of Ownership

People sometimes add hardware to solve performance problems. If a DI tool is strong enough, you can reduce the amount of hardware resources required to support powerful processing—plain and simple.

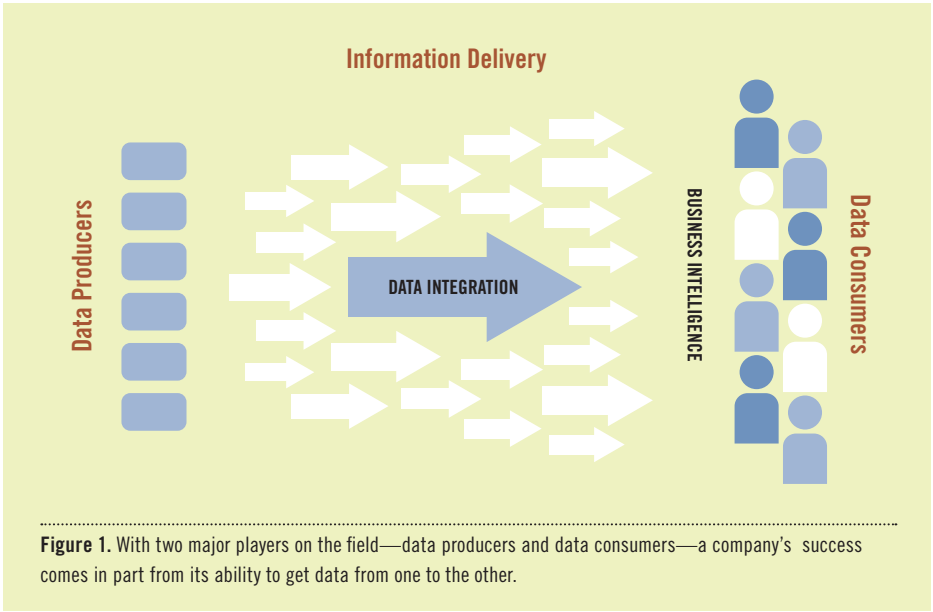


Figure 1. With two major players on the field—data producers and data consumers—a company's success comes in part from its ability to get data from one to the other.

Test the Solution

Before committing to a purchase, test the product in your own environment with your own data. Needs are never the same, and this is the best way to identify the best solution for you. 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.

Value Ease of Use

Particularly in businesses built on hand coding, graphical user interfaces are severely underrated. When it comes down to it, a product that is easy to use can help control the cost and time of training new staff members who are not likely hand-coders.

Real-Life Example

Our customers typically see real BI value in powerful data integration. One direct marketing company saw significant performance improvement before committing to a solution.

The company's primary focus is building and managing customer databases for *Fortune* 1000 corporations. 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.

It took three days to run analytics on 240 GB of data with the tool the company was using. Through a software evaluation in their own environment, the process was completed in less than 10 hours. They determined the solution would give them the ability to run reports more frequently than they were—giving them timely, actionable data.

Conclusion

When it comes to information delivery, it may seem like you are perpetually spinning your wheels. Finding a powerful DI tool to support your BI solution can greatly shift this burden. It also ensures that users across the enterprise can get the actionable data they need—how, where, and when they need it. ●

LESSON FROM THE EXPERTS

Implementing Real-Time Data Integration Solutions

By Jennifer St. Louis
Manager of Marketing, DataMirror, IBM Software Group

With customer loyalty waning and significant competitive pressures on organizations in every industry, companies need to find innovative ways to win business so they can stay ahead of the competition.

Delivering trusted information wherever and whenever needed, in line and in context, to specific people, applications, and processes, allows organizations to gain a competitive advantage in their marketplaces. Companies need to respond more quickly, gain operational efficiencies, and deliver superior customer service by keeping a real-time pulse on their business. Although most organizations already possess the information required to do this, they are often unable to get it when and where they need it.

A real-time data integration strategy ensures accurate data flows across the enterprise so that organizations can make quick decisions on pricing, shelving, service, and product mix, based on the latest information. Managers can analyze critical business data throughout the day to target their marketing efforts, improve up-selling and cross-selling strategies, and better service their customers, no matter what business they're in.

There are four key technical challenges organizations face when implementing a real-time data integration strategy:

1. There is too much information, and users are unable to determine which information is important.

Organizations have great volumes of data and know they can benefit from this information; but many do not know how to pull it together and make it useful to the business. Common examples include not using demand signals to drive supply chains in the retail space or not enabling customers with the most up-to-date banking information in the financials vertical. Companies that do consolidate and transform data can optimize their supply chains, get better results, and improve customer satisfaction.

A solid, real-time data integration strategy can give organizations the information they need to get real insight into their business and create a competitive advantage.

2. Data contains multiple versions of the truth.

Different systems store different pieces of information on core entities, such as product, customer, or partner. This information quickly gets out of synch, making it difficult to get a single view. Commonly, the result is that organizations have problems in managing customer, product, and partner interactions. Also, regulatory compliance is inhibited by poor visibility into the information needed.

3. Customers face issues associated with a lack of understanding of their data.

Without insight into the quality of their data or where the right information resides, organizations may have incomplete and out-of-date or inaccurate data that can quickly proliferate throughout the organization, causing rework, lost time, and lack of trust from end users.

4. Organizations must deal with a lack of agility.

Organizations are unable to take advantage of opportunities for innovation because their systems are too inflexible. They cannot adjust quickly enough to take advantage of new opportunities, and systems are expensive to maintain. Costs tend to escalate due to inflexible systems and manual efforts required to meet the changing needs of the business.

For these reasons, a solid, real-time data integration strategy can give organizations the information they need to get real insight into their business and create a competitive advantage. Yet with growing volumes of data residing in multiple applications, it hasn't always been easy to get trusted information when you need it most.

Implementing a real-time data integration solution

The majority of organizations use ETL tools as their primary means for data integration. They use these tools, or homegrown data integration processes, to extract data in bulk from their production systems and load it into other systems, including data warehouses. The main strengths of these tools are that they extract data from many different applications, perform complex transformations and data quality on that data, and then load large volumes of data into data warehouses. But when it comes to extracting data from production systems or providing real-time visibility into operational systems, ETL tools sometimes need to leverage real-time capabilities. Increasingly, it is becoming necessary for organizations to conduct business around the clock. As more business is done across time zones and over the Web, more organiza-

tions are faced with the problem of shrinking batch windows—making it more difficult for traditional ETL tools to extract data in the short time available. These tools were not built for keeping multiple applications in synch with real-time data feeds.

Implementing a data integration solution should always start with examining your business requirements and deciding where and when you need information delivered. IT then needs to look at where data resides throughout the organization.

This step involves meeting with all interested parties—LOB managers, end users, application programmers, database administrators, etc. During this process, collect all information about the existing IT environment, including what systems the data resided in and where and when the information needs to be accessed. Perform analysis to determine the application transaction volumes to assist in optimizing the implementation.

In the next step—analysis and design—IT must define the project scope and architect the optimum replication scenario for the environment. At this point, complete the analysis and documentation of the project and environment. Recommend replication architecture and solutions, and determine timeframes.

In architecting a solution, keep in mind the effect it will have on source applications. Change data capture solutions can be used to complement or leverage existing ETL processes by providing real-time data flows that capture transactions directly from database logs and send those transactions to existing processes. Take care not to add additional workload on production applications and networks.

After the solution is determined and implemented, completely check the environment to ensure the data integrity and health of existing production applications. The impact of implementing real-time data integration must not negatively affect users of production systems. During this step, review the complex business rules to ensure that they meet the needs of the user environment. Repeat this step to ensure that the solution continues to meet the needs of the organization.

Implementing a data integration solution should always start with examining your business requirements and deciding where and when you need information delivered.

Once a real-time data integration strategy is implemented, it can ensure that accurate data flows across the enterprise, allowing organizations to synchronize information across all their customer touch points. Organizations can gain an immediate single, complete, 360-degree view of customers so they can target their marketing efforts and improve their cross-selling strategies.

Today, only enterprises that can effectively manage, integrate, distribute, and utilize their data assets will survive and prosper. An increasing number realize that without continuous, real-time data integration, they cannot achieve the required visibility into their existing systems. Taking real-time data flows to the next level will enable organizations to sense and respond to data changes in real time. With this ability, they can proactively service their customers and support initiatives, including dynamic warehousing, master data management, SOA, migration/consolidation, and e-business. ●

For a free white paper on this topic, download "Evaluating Real-Time Data Integration Solutions," or click here for more information about IBM.

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, an SAP company

Q What are the key trends affecting business users, and how will Business Objects, an SAP company, empower you to respond?

A Three trends are transforming the business user environment. 1) Companies need to make decisions based on data from inside and outside the enterprise and from structured and unstructured sources. 2) Users need better tools and applications that support collaborative decision making. 3) Companies are seeking competitive advantage by extending their business networks with partners, suppliers, and customers. Today, Business Objects provides the market-leading BI platform and tools that unlock trusted information and enable business insight, performance management, and financial management—independent of the underlying business applications and data stores. While maintaining this portfolio, it will extend its applications offering to help business users, teams, and companies enhance collaboration through networks.

ANALYST VIEWPOINT

Business users are being prodded away from reliance on spreadsheets towards more sophisticated BI tools that better support collaborative decision making and near-real-time interaction with internal and external data and partners. With 44,000 global customers, Business Objects is a leading option for transitioning from spreadsheet environments; the question is how SAP's $6.8 billion acquisition of Business Objects influences the installed base and future adoption. A TDWI survey, "BI Solutions for SAP," found 46 percent of respondents used Business Objects, versus 37 percent for Cognos (IBM) and 30 percent for Hyperion (Oracle). The ingredients are in place for an end-to-end SAP/Business Objects platform, but as with any acquisition, integration will be key.

Collaborative Consulting

Q Do any solution providers have a complete integration platform that would cover all aspects of data integration?

A Data integration has evolved from fundamentally two architectures: batch-based or ETL, and message-based or EAI. ETL vendors have focused on more complex transformations and processing sets of records quickly. EAI vendors have emphasized individual transactions or messages, more simplistic transformations, and real-time data movement. While the tools have begun to converge, there is still a substantial functionality difference between the two technologies. Most ETL providers, while purporting to have real-time capabilities, don't have as robust a messaging architecture as EAI vendors. Conversely, while most EAI solutions have data transformation capabilities, they don't have nearly as complete a set of functions or the volume processing capabilities of ETL. Most complex data integration environments will find a use for both architectures.

ANALYST VIEWPOINT

So far, no mature one-size-fits-all integration platform has emerged. Large enterprises instead have tended to take advantage of capabilities unique to ETL and to EAI with discrete deployments that answer tactical business needs—for instance, real-time EAI capabilities in financial services and the high data volume capacity of ETL for CRM. Both ETL and EAI vendors have made strides in interoperability between the technologies, and some large organizations with complex data integration needs have successfully customized hybrid solutions that capitalize on the best of both worlds. Look for further integration of the integrators, so to speak, as demands for enterprise integration continue to grow.

DataFlux

Q How does data governance drive data integration and MDM projects?

A The policies and practices that form a data governance program provide the core discipline and perspective needed for successful data integration and MDM programs. Any time data is integrated or consolidated, you need a set of uniform policies to guide this process. Data governance techniques and technologies facilitate the creation of these policies, which become the business rules that govern the consistency, accuracy, and reliability of corporate data.

ANALYST VIEWPOINT

Data governance is occasionally paid little more than lip service. That's a recipe for failure. For larger scale data integration and MDM projects, a clearly articulated data governance plan is essential to long-term success. Start small and prepare to grow your data governance program as stakeholders hammer out common data definitions and reconcile data, process, and political issues. A data governance initiative with a set budget, strong executive sponsorship, and tight collaboration between business and IT can and should be in place to help organizations make the most of data integration and MDM.

DATAllegro

Q How can DATAllegro be used to augment my Teradata investment?

A As batch windows for data loads grow smaller and smaller, Teradata customers are running out of bandwidth to perform aggregations of data for business intelligence reporting. DATAllegro provides a suite of utilities to import data from Teradata. The DATAllegro appliance can then be used to run the aggregations and export the data back into Teradata.

A key component of the Teradata Utilities Suite is the ability to directly load Teradata's binary export file format. The atomic-level data is first exported as a binary file from Teradata. Then the data is aggregated into "summary tables" using DATAllegro's high-speed appliance. Finally, the aggregated tables are exported and loaded back into Teradata.

ANALYST VIEWPOINT

DATAllegro fired a salvo in the ongoing data warehousing appliance war with its February 2008 announcement of a new set of utilities to migrate from or augment Teradata systems. A month earlier, a report from Ventana Research recommended that customers considering deployment of data warehouse appliances rather than increasing the number of their Teradata systems should look carefully at the forthcoming Teradata 12.0, with faster batch loading and query performance. For customers, it's all good news—competition is helping to drive down price points, spurring innovation, and increasing choice and flexibility.

Dataupia

Q We have several data warehouses built on various platforms. How do we achieve transparency across our large and diverse data sets?

A Having multiple platforms with more than one architecture is a common situation. One approach is to bring data together and rationalize on one platform. Another is to integrate it before the application layer. Both are costly in terms of hardware acquisition, implementation, and infrastructure maintenance and can be very disruptive.

An alternative is to create a single, but virtual, pool of data where applications can access data, regardless of the source platform, in the format to which they're accustomed. This level of transparency is achieved at the platform's federation layer without disrupting interfaces between applications, database servers, and the physical data.

ANALYST VIEWPOINT

Federation, or enterprise information integration (EII), is an increasingly common way to leverage data not only in multiple and heterogeneous data warehouses, but in data marts, relational databases, and applications as well. Federation offers the advantage of quick hit successes and enabling business users to query independent data sources without the data actually being moved. On the downside, this "loosely coupled" data warehouse architecture is not well suited for computationally intensive queries. For some organizations, a hybrid approach that utilizes both federation and a single enterprise data warehouse will deliver the greatest bang for the buck.

IBM

Q What impact does real-time reporting have on existing systems?

A The impact depends on the business need. For transactional level reporting, minor changes may have to be made to the warehouse schema to support transactional details from operating systems. If the need is for frequent reporting on aggregates, the frequency and its impact on ETL jobs will have to be determined. The amount of data may have an impact on operating systems if more frequent extracts are required. Understanding the impact will help make the decision of leveraging changed data capture technologies that can mitigate the risk to production environments.

ANALYST VIEWPOINT

The demand for real-time information has given rise to operational BI and introduced vexing questions that organizations must address. The first technical question that an organization needs to address is whether to use a data warehousing architecture to deliver just-in-time data or bypass it altogether. Without a warehouse, organizations need to be careful to avoid generating conflicting data sets. Operational BI alternatives to the warehouse approach include federated query models, event-driven analytic engines, integration with a message bus, and other techniques. Interestingly, a TDWI survey found 51 percent of respondents running both operational and analytic BI in the same environment.

Identity Systems

Q What is the most effective and efficient strategy for finding duplicates in a large customer database?

A The most successful way of finding a name match in a database is, first, to perform a search on an index built from name alone, thus building a candidate list of possible matches. Then refine, rank, or select the matches in that candidate list, based on other identification data.

The more of the name used in the key, and the greater the number of keys built per name, the greater the variety of search/match strategies that can be supported.

In large-scale systems, the choice and sophistication of the search/match strategy is consequential to performance demands, risk of missing critical data, need to avoid duplication of data, and the volume of data under indexing.

ANALYST VIEWPOINT

This approach requires that the database have a customer name index. Luckily, database administrators usually create such an index in databases that manage customer data. However, it might be preferable to create a specialized index just for customer lookups and matches, to give these faster performance. Some vendor products for data quality or identity searching can automatically create and maintain specialized indices which—unlike an index in a standard database—can cope with the misspellings, typos, and other quality problems typical of customer data.

Information Builders

Q What's the basic difference between analytical BI systems and operational BI systems?

A Analytical BI systems generally access a data warehouse. They give users an excellent view of past business events and entities but not of current business processes, which are ongoing. Analytical BI applications rely on extract, transform, and load (ETL) tools to keep a data warehouse current, perhaps once a day or once a week. Operational business intelligence systems, by contrast, give users a real-time view of business events as they occur. These BI applications generally obtain information from an automated workflow process or directly from production systems.

ANALYST VIEWPOINT

Another notable difference is the application and functional areas to which analytic BI and operational BI are applied. So far, operational BI has had the greatest appeal for time-sensitive, mission-critical systems such as fraud detection and supply chain, with negligible value for, say, HR. Operational BI adoption is steadily growing. A TDWI survey in mid-2007 found that 53 percent of respondents reported that their organizations were doing operational BI with intraday data delivery, though only 16 percent characterized their implementations as mature. Expect both adoption and maturity to increase as organizations take advantage of the real-time capabilities of operational BI.

Initiate Systems

Q What should I look for when selecting an MDM vendor?

A Look for a vendor focused on MDM. Your needs will evolve over time, so you need a vendor that will grow with you and whose product roadmap won't get diluted by other products.

Look for a vendor that develops its core capabilities. As your needs evolve, you need a vendor that has control over its roadmap to ensure your future needs are supported.

Finally, look for a vendor focused on quick time to value. The longer your project incurs costs with no benefits, the less likely it is that you will recover your investment or obtain future executive support.

ANALYST VIEWPOINT

Organizations will do well to assess both how well the solution addresses today's tactical challenges and aligns with longer-term strategic needs. As MDM is fundamentally a data integration practice, assess the compatibility of an MDM solution with your existing ETL, EII, EAI, or other integration infrastructure. Look for cost efficiencies in repurposing your data integration technologies and expertise. From a strategic perspective, recognize that MDM is still in its early phases and is likely to evolve beyond customers and products to include such entities as employees, suppliers, and partners. Ensuring an MDM vendor's roadmap aligns with your strategic objectives is rarely easy but is a key factor in realizing long-term value.

MicroStrategy

Q Why use advanced data visualizations to display business intelligence information?

A Advanced visualizations express data in more meaningful ways than is possible with traditional grids and graphs. Two important facets in visualizations improve data comprehension:

Information presentation. The size, color, and animation of visualizations are based on the data, making it possible to quickly spot exceptions and anomalies.

Information density. Large reports are collapsed into a single visualization, providing a bird's-eye view of the entire data landscape without the need to scroll through the data.

Advanced visualizations let users make more informed decisions by providing timely, relevant, and accurate information to answer their business questions.

ANALYST VIEWPOINT

Fundamentally, data visualization is nothing new. But in the past year or two, this field has hit its stride as vendors successfully married rich visualization capabilities with practical tools and technologies (e.g., dashboards and event processing). The best data visualization solutions excel at both style and substance to give users an engaging visual medium atop a strong analytics platform with deep reach into disparate data, ranging from desktop sources to data warehouses. Done right, data visualization can encourage strong user adoption and deliver genuine business insights by transforming slate-gray numeric data into a lively visual environment that highlights outliers and enables drill-through to generate business answers.

SenSage

Q What is the advantage of using a different data warehouse solution for event data over my current solution?

A While your current data warehouse solution is more than capable of storing event data, the biggest advantage is cost. Performance improvements are worth noting, but most executives are more interested in capital expenditures.

Since the SenSage Event Data Warehouse uses a columnar database for storage, event data is compressed at a very high rate. This reduces the storage requirements by an order of magnitude and therefore reduces storage capital expenditures. Additionally, the CPU required to search the reduced amount of data further reduces costs. Finally, software licensing from SenSage is typically less than traditional data warehouse solutions while containing ETL and analytic tools required for event data.

ANALYST VIEWPOINT

A dedicated event data warehouse can make sense for large organizations with a need to closely manage and analyze log files and other event data to strengthen security and regulatory compliance. Banking, telco, payment processing, and healthcare companies are among those increasingly deploying security information and event management (SIEM) solutions and dedicated event warehouses to manage terabytes of event data, including log information generated by network equipment. Prospective buyers should recognize that SIEM solutions can be complex and be prepared to scrutinize an array of solutions in this rapidly expanding sector before selecting one that best meets their needs.

Siperian

Q What's the difference between technology-focused and business-focused MDM starts?

A Technology-focused MDM starts advocate that companies start with a single data type (such as customer), implement MDM using a small footprint (such as registry style), or deploy MDM solely with a data warehouse to improve reporting. These approaches may limit the scope and potential return on investment (ROI) from MDM, since they do not attempt to solve the most pressing and difficult business problems. MDM is more precisely about solving business problems by efficiently managing master data that is critical to a company's business operations. Consequently, a business-focused approach can provide a complete MDM solution that addresses the specific business problem and provides tangible business value and significant ROI in a short-term timeframe.

ANALYST VIEWPOINT

A technology-focused approach to MDM can make sense when an organization is suddenly confronted by large volumes of inconsistent and redundant customer, product, or other master data that compromises business performance. A merger or acquisition, for instance, can result in master data influx of crisis proportions. What's important is that any tactical effort to attack the problem at a technology level aligns with a broader, business-focused MDM strategy. Technology- and business-focused MDM can and should evolve in lockstep in an incremental, multi-year evolution towards the common goal of a single, trusted set of master data that delivers measurable business value.

Syncsort

Q How can I speed querying in a data warehouse?

A Aggregates are the best way to speed warehouse queries. A query answered from base-level data can take hours and involve millions of data records and millions of calculations. With precalculated aggregates, the same query can be answered in seconds with just a few records and calculations.

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

ANALYST VIEWPOINT

With data volumes growing inexorably, data warehouse query performance is becoming a significant issue that can derail BI effectiveness. Increasing user populations, complex and concurrent queries, and demand for near-real-time information complicate the challenge. Organizations in a performance pinch need to closely examine root causes in their DW environment and select the solution, or combination of solutions, that best addresses their situation. Six common approaches are (1) brute force (more hardware and software licenses), (2) incremental tuning, (3) migrate to a new DW platform, (4) use memory caching or DW appliances, (5) aggregate data subsets for rapid retrieval, and (6) index data to accelerate retrieval from base tables.

Talend

Q What are the main benefits of open source data integration?

A Open source data integration has matured and is now technically on par with, or superior to, traditional, proprietary solutions. Open source brings to the table a different business model. Open source requires no initial investment so projects can get started easily, and carries no per-source/target or per-CPU costs so deployments are not restricted by funding issues. Users only pay for the support they use. From the openness perspective, it reduces the dependency of the user, and additional connectors and features can be built easily into the product. However, open source is not free. Technical support, training, and costs of development must be considered, but it's a lot less expensive than the alternatives.

ANALYST VIEWPOINT

Open source data integration software is proving attractive for departmental or tactical deployments, resource-constrained government, educational, and nonprofit organizations, and ISVs and Web 2.0 players using LAMP (Linux, Apache, MySQL, PHP). From a cost perspective, it's hard to beat free downloads, low-cost support, and a freely available community knowledge base. While it's steadily maturing, open source is still a ways from matching the big data integration vendors on such capabilities as integrated data profiling and quality, changed data capture, high availability, and robust metadata management. Most organizations running data integration in mission-critical systems will stick with the status quo for now.

Data Governance Strategies

Helping Your Organization Comply, Transform, and Integrate

BY PHILIP RUSSOM

Definitions of Data Governance

Data governance is hard to define because it's still new and evolving. Each organization tailors data governance to its needs and abilities, and DG is practiced both in isolated pockets as well as on an enterprise scale. Furthermore, DG is inherently a cross-functional program that involves a mix of technology and business people—plus their IT systems and business processes—and the mix varies greatly.

Even so, here's a definition that covers almost all the components and goals of data governance:

> Data governance (DG) is usually manifested as an executive-level data governance board, committee, or other organizational structure that creates and enforces policies and procedures for the business use and technical management of data across the entire organization. Common goals of data governance are to improve data's quality; remediate its inconsistencies; share it broadly; leverage its aggregate for competitive advantage; manage change relative to data usage; and comply with internal and external regulations and standards for data usage. In a nutshell, data governance is an organizational structure that oversees the broad use and usability of data as an enterprise asset.

That's a mouthful. So, here's a rule of thumb that's easy to remember:

> DG usually boils down to some form of control for data and its usage.

The catch is that "control" has multiple meanings that are somewhat at odds:

- **DG may tighten control to limit data access.** This is true when data governance is driven mostly by compliance goals, especially data security and privacy.
- **DG may ease control to expand data integration.** Most DG boards provide procedures through which a team can request access to data owned by another team. Ironically, this eases the control of data to assist initiatives that rely on broadly integrated data, like business intelligence (BI) and customer relationship management (CRM).
- **DG may define controls that improve the content of data or dictate its structure.** For example, data flows through many IT systems and departments, so improving the quality of data (whether physical or semantic) is a cross-departmental affair that DG can manage. Likewise, enterprise data architecture seeks to tweak the structure of multiple databases for the sake of easier database management or data integration. DG can define standards that dictate consistency for data structures and data definitions.
- **The level of control can vary.** For example, strict governance is typical of federally mandated compliance, whereas loose guidance is typical of data architecture standards. Or, a multi-divisional corporation may demand strict governance for data at the headquarters level so data yields a unified view of total corporate performance, yet merely provide loose guidance for individual implementations so local organizations can satisfy local requirements.

"In our consulting practice, we have participated in data governance initiatives that evolved from either grass-roots data management or executive fiat," said David Loshin, president of consultancy Knowledge Integrity, Inc. "In one situation, the need for standardizing shared data representations drove the 'bottom-up' development of a governance infrastructure, leading to a federated data standards governance framework. In another situation, the introduction of a consolidated enterprise application suite was expected to be accompanied by data governance as directed 'top-down' by senior management. Whether bottom-up or top-down, both cases posed common challenges in communication, standardization of concepts, and establishing operational processes for governance. On the technology side, data quality, metadata, and policy management were success factors."

Critical Attributes of Data Governance

These definitions help us understand the goals and actions of most data governance programs. To fill out the rest of the picture, here are other attributes of successful programs. Note that all are core assumptions of this report, and all are discussed in detail later:

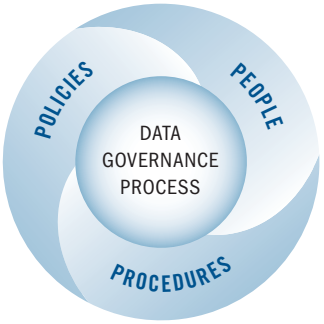


Figure 1. The data governance process consists of people, procedures, and policies.

- **Data governance is mostly about the "four Ps."** These are seen most clearly in the DG board, where people work together to establish and enforce policies (or rules) defining which data is subject to governance, as well as the allowable access and usage of such data. Procedures provide a structure for reviewing and acting on requests for data access, data improvement, and other changes. People, policies, and procedures all combine to enable a larger DG process (see Figure 1). The four Ps explicitly remind us that DG is mostly about people collaborating to establish a DG process that accommodates the needs of all their business units (and external entities, when appropriate), but with priority to enterprise goals.

- **Data governance must coordinate with other forms of governance.** Don't forget that data governance is but one form of governance. TDWI Research has interviewed people who've made DG work in isolation (say, just for BI or data quality). But, in the long run, DG should coordinate with other forms, especially IT governance and corporate governance.

- **Data governance doesn't govern data directly.** The term "data governance" leads us to believe that we are governing data directly. But the truth is that we're governing how data is accessed and used via business initiatives, as well as defined and managed via data management infrastructure. This explains why DG is increasingly a component of these initiatives and infrastructures.

- **Data governance intersects with business initiatives.** An assumption of this report is that data governance touches many different business initiatives, especially those that are data-driven, like compliance, BI, CRM, and business transformations. DG is often a subset of these initiatives, and is increasingly a critical success factor for them.

- **Data governance intersects with data management practices.** When executed broadly, DG influences almost all data management practices, including data quality, integration, warehousing, standards, administration, architecture, and lifecycle management. DG typically requires that adjustments be made in these practices, in support of the policies developed by the DG board. While tools dedicated to DG are rare today, some data management tools can automate some actions of DG.

- **A successful DG program strikes a pragmatic balance among competing goals.** For example, there's a prominent need for balance between compliance goals that limit data access and business integration goals that expand data access. Other opposing goals include business versus technology, data content versus data usage, strict governance versus loose guidance, and departmental versus enterprise data ownership. Most DG programs start in one carefully bounded area that serves a single goal (like DG just for BI, compliance, or master data management [MDM]), so the balancing act is not immediately apparent. Striking an appropriate balance becomes a critical success factor as the program expands to govern more data sets, data usage scenarios, and data management practices. Such balances are difficult to attain and maintain

without the executive mandate, central policy making, change management procedures, and cross-functional collaboration of data governance.

Why Data Governance Now?

There are many reasons why organizations should initiate or expand DG programs now:

- **The current "age of accountability" demands compliance.** And punishments for non-compliance are severe, ranging from customer flight and revenue loss to fines and jail terms. Firms are under unprecedented pressure to control data usage according to internal policies for data security and privacy, as well as external regulations like Basel II, HIPAA, and SOX. Assuring compliance is an early-phase goal of most DG programs.[1]

- **Compliance and business intelligence demand high-quality, auditable data.** Organizations need to improve the quality of data that goes into public documents, especially regulatory reports. Furthermore, report auditability—i.e., recording the lineage of report data—is crucial to surviving an audit, regardless of who the auditors are. And one of the most common questions asked by report consumers internally is: "Where did this data come from?" Today, the quality of report data is a high priority for most DG programs, whereas auditability is a lesser priority.

- **Improving data quality is a cross-functional imperative.** Since a DG board is cross-functional by nature, it's an ideal organizational structure to effect improvements that span multiple business units. Although data quality focuses mostly on physical data, master data and metadata need improvement, too. This is why many data quality and master data management initiatives are supported by a cross-functional DG board.

- **Data integration (DI) implementations cast an ever-widening net.** This is true whether DI is analytic (feeding a data warehouse), operational (consolidating database instances) or cross-business (sharing data with partners). DG can both limit these implementations to assure compliance and liberate them to reach more data sources and targets. DG can also assist by providing data exchange standards and procedures for data access and improvement requests.

- **Data governance reduces the risk incurred during business transformations.** DG is imperative in firms that experience regular transformations such as reorganizations, mergers and acquisitions, and initiatives that involve data as an enterprise asset (typically linked to CRM or sometimes BI). These transformations require extensive changes in data ownership and data structure. DG can manage the changes while assuring compliance. ●

Philip Russom is the senior manager of TDWI Research at The Data Warehousing Institute, 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, 30-page report by the same name. You can download this and other TDWI Research free of charge at www.tdwi.org/research/reportseries.

The report was sponsored by Business Objects, an SAP company, Collaborative Consulting, DataFlux, Exeros, Informatica Corporation, SAP, SAS, and Trillium Software.

1. Visit TDWI's White Paper Library (www.tdwi.org/WP) to download the IT Audit Checklist Series, which goes in depth into several issues mentioned in this report, including information security, data privacy, IT governance, and change management.

CASE STUDY

American Heart Association Builds a Unified View of the Customer

Commentary by Jon Gerush
Manager, Data Integration and Management, American Heart Association

The Business

The American Heart Association is a national voluntary health agency whose mission is to build healthier lives, free of cardiovascular diseases and stroke. The nonprofit organization is committed to fighting heart disease and stroke as well as raising awareness of these diseases. As part of its mission, the organization focuses on specific causes designed to help people achieve a heart-healthy lifestyle.

The Challenge

After years of adding new systems to track customer information (including healthcare professionals, donors, committee members, and newsletter subscribers), the American Heart Association faced a customer data integration (CDI) problem. The IT infrastructure included multiple customer data stores for different business units.

With multiple applications containing more than 14 million names, often housed in disparate, disconnected systems, the organization had no reliable way to manage donors across these systems. The staff had difficulty reconciling these disparate views or understanding an individual's true value across the organization. With this effort, they could effectively work toward a centralized repository of customer information that gave the organization the ability to put quality data into the hands of business users.

While the initial data quality project focused on standardizing addresses and identifying duplicates rather than a classic CDI solution, the company expected that any application selected should contribute to a future CDI implementation. The selected application needed to stand on its own and to have proven value within and outside of CDI initiatives. The solution needed to be external to the enterprise applications (i.e., CRM) so that it could be leveraged to improve data quality for several internal and external data sources.

The DataFlux Solution

The American Heart Association chose dfPower Studio and the DataFlux Integration Server to meet its needs. dfPower Studio allowed business users to create advanced matching rules using the system's advanced fuzzy matching capabilities, seeking out duplicate records and consolidating customer data into a single master record. The DataFlux Integration Server provides the ability to extend rules and policies created in dfPower Studio throughout the enterprise.

dfPower Studio helped enhance the value of contact information through standardizing and validating postal information, assigning gender to customer records, and moving information such as prefix, suffix, and lineage into the proper fields. These rules provide the foundation for CDI efforts, providing a single repository of data quality processes.

The Results

With DataFlux solutions in place, the American Heart Association is transforming its customer information into high quality data.By deduplicating data across systems, the organization had a real foundation for more in-depth master data management (MDM) efforts.

"The dfPower Studio interface is easy to use and intuitive," says Jon Gerush, manager of data integration and management for the American Heart Association. "This certainly helped our staff understand data quality issues during profiling and allowed for the fast implementation of data quality and data integration jobs, which could then be embedded into other jobs."

DataFlux technology is alleviating the organization's duplicate customer data issues. The American Heart Association now manages more than 12 million customers in its Siebel systems with less than five percent overall duplicates. With better data, the organization has begun to expose processes to other data stores, integrating these multiple databases to create a foundation for a true CDI initiative. For the American Heart Association, the decision to take a phased approach to MDM provided real benefits, as it allowed the organization to create a data quality and data integration framework that could drive both current and future data management efforts.

"DataFlux technology provided a method to analyze, improve, and control our vital donor information. From a services and support standpoint, DataFlux always made the organization feel special," says Gerush. "DataFlux staff was quick to respond to any inquiries through a combination of professional services, online support, and the customer portal." ●

For a free white paper on this topic, download "Five Steps to More Valuable Enterprise Data," or click here for more information about DataFlux.

LESSON FROM THE EXPERTS

Adding Data Governance to Improve Business Decisions

By Tony Fisher
President and CEO, DataFlux

Anyone who has ever flown a plane—or even glanced into a cockpit when boarding a commercial flight—can appreciate the complex array of gauges and monitors that the pilot must check. All the data about a plane's speed, course, fuel, and other details are within easy sight, each giving the pilot the information necessary to make sound, safe decisions.

Similarly, organizations rely on data to provide the foundation for business decisions. For years, companies have implemented business intelligence (BI) programs to achieve one goal: make better decisions from their corporate information. Many companies have discovered one inescapable truth: it's impossible to make an informed decision based on outdated or erroneous information. Just as a pilot needs to monitor the health of the aircraft, organizations need to constantly gauge the health of their data.

"Once and Done" Is Not Enough

The impact of data decay can influence—and hinder—many enterprise initiatives. Imagine a manufacturing company that builds a data warehouse to serve as a single repository for all of its information about customers, products, and inventory. From that data, it can uncover trends about customer adoption, resource allocation, and future needs.

After a review of the data, this company finds that new, nonstandard information is constantly arriving at the repository. The effect of this bad data may not be felt until much later. Whenever the company explores this data to identify patterns or tendencies, the presence of bad data can skew the results.

The solution for building high-quality corporate data on an ongoing basis is data governance. With data governance, technology and business users can create rules to examine data automatically to uncover problems as they occur. These users can also chart metrics related to data quality on a periodic basis and begin to address some of the underlying reasons that bad data is being collected in the first place.

The Role of Data Monitoring

Just as pilots continually monitor their gauges, companies need to steward their data as a valuable resource. Instead of loading questionable information into their data warehouses, companies can use data governance programs, which often include data monitoring, to check and control incoming information in order to maintain high levels of data quality.

With data monitoring, you can:

- Detect problems from incoming data. Validate existing data against established business rules to uncover and address data integrity issues—before they become a problem.
- Generate instant alerts. Set up automated system notifications and e-mails to flag problematic data as a new, inconsistent record enters the system.
- Identify trends in data quality metrics. View ongoing statistics about data to see when the value of data starts to decline.

Data monitoring extends the reach of traditional data quality programs by making good data a corporate priority. When data does get out of control, users know immediately—and they can react to problems before the quality of the data declines.

For organizations that have already started an effort to improve data quality, most of the elements are already in place to build a data governance program. In fact, data monitoring is an extension of the effort required to get data into a reliable state in the first place. The same business rules used to cleanse, standardize, and verify data in the initial data quality project can serve as the rules to examine and flag data integrity issues over time.

Building consistent, accurate, and reliable data is not easy. Periodic fixes will only provide temporary relief from the various problems that can arise because of bad data. With data monitoring, companies can better control their data and build more reliable information to support any future business intelligence efforts. ●

CASE STUDY

Subex Deploys Data Warehouse Appliance to Power its 150 Terabyte OSS Systems

Commentary by Samatha Stone
Vice President, Marketing, Dataupia

Subex Limited is a leading global provider of Operations Support Systems (OSS) that empowers communications service providers to achieve competitive advantage and deliver new service experiences to subscribers. Subex's customers include 32 of the world's 50 largest service providers. The company has more than 150 installations across 60 countries.

The company pioneered the strategic concept of the revenue operations center—a centralized framework for end-to-end control of a service provider's revenue costs, fostering operational dexterity for sustained profitability.

When Subex sought to replace one of its legacy data warehouse systems, which could no longer accommodate growing data volumes, they chose the Dataupia Satori Server. Dataupia's data management system operates as the supporting framework for one of Subex's 150 terabyte systems, enabling increased scalability to manage call detail record (CDR) volume growth and support enhanced business intelligence.

The Dataupia Satori Server improves productivity at Subex by increasing concurrent system usage. One Subex business unit immediately noticed a dramatic increase in load speed, query performance, and diversity of the types of queries now able to run. The Subex can now load more than 1 billion CDRs per hour. Complex queries that would previously have taken two weeks are now possible in a few hours.

The storage and retrieval of CDRs is vital for supporting billing disputes and reconciliations, financial analysis, network optimization and revenue assurance. CDR data is used to make critical business decisions, assist margin analysis, and provide a single consistent view of product and customer behavior. The Dataupia Satori Server provides Subex with broader access to CDR data by retaining large amounts of data online that can be accessed by reporting or analytics applications.

"Dataupia has helped us do things with our corporate IT infrastructure that we have never been able to do before," says Paul Skillen, president, BT Business Unit, Subex Ltd. "With the Dataupia Satori Server we were able to power complex analytic requests that were not possible before—query performance on a large scale is simply outstanding."

"Dataupia provides us with an economical solution to help maintain that focus by providing more access to data. The ability of the Dataupia Satori Server to handle large volumes of data across concurrent users helps increase the value we obtain from that data, as well as the overall productivity of these providers."

Paul Skillen, Subex Limited

Key drivers leading to the implementation of the Dataupia Satori Server included scalability, ease-of-use, cost effectiveness, increased productivity through concurrent usage, and environmental benefits such as lower power consumption. Additionally, Dataupia's omniversal transparency allows Subex to experience seamless integration with existing systems resulting in an easy deployment process.

"There has been a shift in the telecommunications industry causing telecommunications providers to see their margins dramatically decrease. In order to remain effective, they need to focus heavily on service agility and operational efficiency," continues Skillen. "Dataupia provides us with an economical solution to help maintain that focus by providing more access to data. The ability of the Dataupia Satori Server to handle large volumes of data across concurrent users helps increase the value we obtain from that data, as well as the overall productivity of these providers."

The Dataupia Satori Server data management system is an all-in-one solution—server, storage, and optimization software packaged as a single appliance—designed specifically to deliver persistent access to as much data as an organization needs. The combination of highly specialized software and powerful processors allows large amounts of data to remain on-line and ready for use. The Dataupia Satori Server installs quickly, requires little administration, and allows for continuous and seamless scalability for increased users and data.

For more information please visit www.subexworld.com. ●

For a free white paper on this topic, download "The Business Value of Having the Right Data at the Right Time," or click here for more information about Dataupia.

LESSON FROM THE EXPERTS

Keep it Simple: Gaining Efficiency Through Data Warehouse Appliances

By Foster D. Hinshaw
CEO, Dataupia

The data warehouse appliance was introduced specifically to address the needs of the "big data" vanguards. Just a few years ago, multi-terabyte data warehouses were rare. Now, almost everywhere you look, organizations have to, want to, or plan to capture, retain, and eventually use vast amounts of data.

The appliance approach relieved IT of having to build its own infrastructure out of a mix of iron, wiring, and hand-coded software modules. If the infrastructure could be streamlined, more time could be spent on the data and information side of the equation—that is, the parts that brought tangible value to the business.

The data warehouse appliance achieved that by packaging together everything needed to build a data warehouse. Its goal was to deliver a "data warehouse in a box." Data warehouse appliances are now far simpler to install and maintain than a typical database server plus storage infrastructure. They're easier to get up and running, but is that enough? How hard is it to "personalize" them so that they can adjust easily to constantly changing requirements?

Personalizing Infrastructure

Why talk about personalization in the context of data warehousing? Why add personalization to the list of desired characteristics for a data warehouse appliance? I believe personalization is the element that brings the data warehouse appliance to the next level of usefulness—or relevance to the business drivers that are behind current data warehouse funding.

By personalization, I mean continually adjusting aspects of the appliance to suit an organization's needs. The market will soon demand the ability to personalize infrastructure as customers start expecting their data warehouses to contribute to their agility instead of defining their limits. Greater flexibility will translate into more demands for scalability, accommodating both complex analytics and routine reporting with everything in between, handling more users wanting to do a greater variety of things with more data.

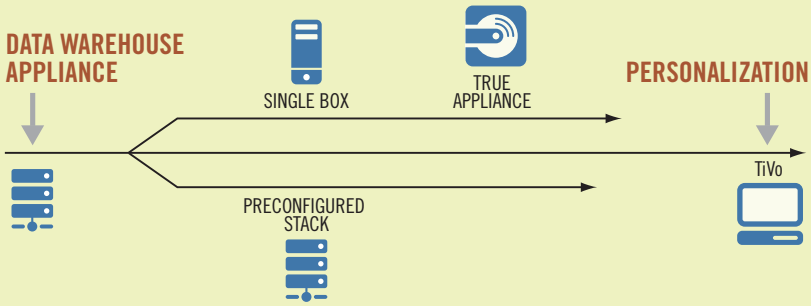


Figure 1. Data warehouse infrastructure architectures can be evaluated as to the degree and ease of personalization, i.e., flexibility, they offer.

How Adaptive Is Your Infrastructure?

To profile your infrastructure's flexibility, consider all the components that have to be touched when you make significant additions to users, data volume, or dependent applications:

1. How many physical parts are involved? Include hardware, cables, power packs, etc.
2. How many changes have to be made to the physical environment? Think green in terms of power, cooling, and space.
3. How many changes to connected hardware are involved? Include all back-up, disaster recovery, and additional storage hardware.
4. How many changes to dependent software—for example, the software used to support items covered in question 3?
5. How much work is expected to get it to meet your basic operating requirements? Which skills, how much staff, and what dependencies on business users' input are involved?
6. How much work will it take to test? Consider IT hours and hours on the business side spent on testing.

Data warehouse infrastructure architectures can be evaluated as to the degree and ease of personalization, i.e. flexibility, they offer.

The higher the numbers in your answers, the more work, complexity, and risk are involved with every change. Traditional tiered storage architectures have very high numbers with their many moving parts and touch points.

The newest data warehouse appliances will have the lowest numbers because they are designed to be non-disruptive, requiring no adjustments to applications and no disturbance in the business user's world. From the IT perspective, they quickly become good "data center citizens," scaling easily, requiring little attention and few resources. They reduce the impact of change, making change feasible and personalization possible.

To build flexibility into your data warehouse, you need to start streamlining each of these change effects. The more minimal the impact of a change, the more easily you can adapt to the next wave of business requirements. Taking the steps you'll need to bring your infrastructure forward to the point where personalizing is business as usual, is now an affordable option you can't afford to ignore. ●

LESSON FROM THE EXPERTS

Maintaining Mixed-Workload Performance While Loading Data

By Jesse M. Fountain
Vice President, Pre-Sales Services, DATAllegro

As data warehouses become increasingly mainstream and mission critical, data warehouse administrators find themselves plagued with the following dilemma: How to keep data refreshed in a near-real-time manner without degrading overall query performance.

Historically, the cardinal rule of data warehouses has been to avoid updating the warehouse during traditional business hours. In recent years, thanks in part to the Internet marketplace, business hours are no longer 9-to-5 but run around the clock. Gone is the quiet window wherein batch data inserts/updates/deletes can take place.

DATAllegro was the only vendor that fulfilled the simultaneous data load requirement.

The most popular workaround has been to maintain two separate instances of the database using a "ping-pong" strategy, with one instance being devoted to queries while the other accepts data loads. Periodically, say every hour, the load database becomes the query database and what was the query database becomes the load database. This strategy is not only expensive—having to keep two instances of the database—but also requires unnecessary ETL complexity.

In a recent proof of concept (POC), a large manufacturing company that uses Teradata required participating vendors to demonstrate their ability to allow simultaneous data loads without affecting concurrently running queries. On a single-rack system, DATAllegro set up the environment to run 200 concurrent users while injecting 50 million rows into the same table being queried. The stand-alone load job took approximately 26 seconds. However, in a concurrent scenario, the job ran in under 60 seconds with almost no detectable (< 10%) degradation to the performance of concurrent queries. At the conclusion of the POC, DATAllegro was the only vendor that fulfilled the simultaneous data load requirement to the manufacturing company's satisfaction.

The secret to DATAllegro's capability lies in the manner in which we load data. First, our high-speed bulkloader capitalizes on speedy MPP architecture to move data onto the appliance in the form of a temporary table at blindingly fast speeds approaching 1.2TB per hour. Next, a standard transaction-safe "select into" is invoked to insert/update the data into the permanent table. DATAllegro's high-speed FASTINSERT facility controls table locks in a manner that does not interfere with pending queries. This entire operation is handled automatically by the DATAllegro Autoloader and is completely failsafe.

Now, organizations with real-time and near-real-time data requirements can rest easy that their end users will not suffer poor query performance while enjoying the business benefits of up-to-the-minute data refreshes. ●

For a free white paper on this topic, download "Transitioning your Centralized EDW to a Hub-and-Spoke Architecture," or click here for more information about DATAllegro.

LESSON FROM THE EXPERTS

Evaluating Data Warehouse Appliances Based on Cost-per-Statement

By Jesse M. Fountain
Vice President, Pre-Sales Services, DATAllegro

As more and more organizations turn to data warehouse appliances (DWAs) to control the spiraling costs of their VLDB (very large database) data warehouse implementations, they most often look for easier ways to evaluate vendors.

As little as two years ago, pioneering organizations were sold on the price/performance of DWAs but focused mainly on single-run query speeds as the primary justification for the investment. Not considered were factors related to system throughput for 200+ concurrent users, let alone throughput for mixed workload queries while simultaneously loading and updating large tables.

Consider that your data loads (time and volume), your database design, the complexity (or simplicity) of your queries, your concurrency/mixed workload, etc., are very specific to your environment and culture. What you really need is an easy way to incorporate these nuances into the evaluation criteria so that you can effectively compare and contrast vendors.

The evaluation criteria for DWAs has matured to include query speeds, data load speeds, concurrency, and backup/restore time. However, some organizations are adding another measure to the mix that provides a more common denominator to evaluate DW appliances.

This new yardstick is often referred to as the cost-per-statement, or CPS. CPS is a more accurate measurement, as it represents not only queries, but DML and load statements as well. It can also include other operations such as system backup and data export.

Here is the simplified equation:

$$CPS = \frac{Annualized\ System\ Throughput\ (Queries + DML\ Statements + Load\ Statements)}{Annualized\ Appliance\ Costs}$$

To build your own customized CPS model:

1. Determine your expected physical system requirements and costs. These costs should include the cost of the hardware, database, operating expenses, implementation/training costs, etc.

2. Determine statements-per-hour (SPH): Select 20–30 of your most "representative." There are various techniques for identifying these, but essentially a query stream from your busiest workday can be pulled and analyzed.

   a. Set up a concurrency model of the representative queries where each query is parameterized with varying values (dates, geographies, etc.). The concurrency model should simulate the number of threads needed to handle your particular workload of users.

   b. A separate model should be set up to represent your particular data loading requirements. In addition to data loads, this model should include DML statements (CTAS, updates, deletes, etc.).

   c. Run each of the models over a period of time and capture average timings to calculate the SPH. Note that if real-time or near-real-time data loading/updating is required, then both models should be run simultaneously.

3. Extrapolate SPH into statements-per-year (SPY), taking into consideration your particular operating window. Example: for a service level agreement where the system is available seven days a week, 24 hours per day, and queries and DML statements run concurrently, the factor is (24 x 7 x 365 = 61,320).

Table 1 is a simple CPS analysis of a 30TB DATAllegro appliance for a retail organization running intensive market analysis queries, simultaneously loading 50 million rows per hour and applying updates to two five-billion–row fact tables.

Somewhat akin to miles per gallon for automobiles, CPS allows you to compare the price tag of various data warehousing infrastructure solutions. While CPS takes into consideration your particular environment to determine a common factor that can be used across all vendors, it does not consider the predictability of run times (minimum and maximum query execution times) in a mixed workload environment. Therefore, CPS should not be the only factor used to evaluate DWAs. Nonetheless, CPS should be part of an evaluation—and, of course, your mileage may vary. ●

	Rate	Year 1	Year 2	Year 3	Total
30TB Appliance	$ 1,500,000	$500,000	$500,000	$500,000	$1,500,000
Hardware/Software Maintenance	Incl	$270,000	$270,000	$270,000	$810,000
Backup System	Incl				
Power Requirements	(4) 30AMP Circuits @ 800/Month ea	$38,400	$38,400	$38,400	$115,200
Cost per Sq Ft (Cooling & Space)	16 Sq Ft X $50/Month	$9,600	$9,600	$9,600	$28,800
Professional Services	30 Days	$60,000	0	0	$60,000
Implementation Conversion Costs	$30/hr	$60,000	0	0	$60,000
Data Center System Admin (½ FTE)	100,000/yr	$50,000	$50,000	$50,000	$150,000
DBA Support (½ FTE)	175,000/yr	$87,500	$87,500	$87,500	$262,500
	Total Costs:	$1,075,500	$955,500	$955,500	$2,986,500

Average Cost per Year	$995,500
Statements per Hour	20,000
Statements per Year	1,226,400,000
Cost per Statement	$0.0008

Table 1. A simple CPS analysis of a 30TB DATAllegro appliance.

CASE STUDY

Using Technology to Support a Mobile Workforce

By Len Nicolas
CIO, Nygård International

Nygård International is a leading fashion company that designs and markets a wide range of women's fashion apparel and accessories. This multimillion dollar fashion business, headquartered in Winnipeg, Canada, sells directly to customers through major retailers and its own retail operation and e-commerce site. It has more than 2,600 employees internationally, with sales offices located in New York, Toronto, and Montreal.

The entire Nygård culture is dedicated to efficiency through technology. According to founder and CEO Peter Nygård, "handling paper is expensive and time consuming, especially when the use of electronic commerce is so much more transparent and immediate."

Nygård supplies apparel to a broad range of retail clientele, including Dillard's, Sears Canada, and HBC. Millions of dollars worth of orders are placed daily and must be shipped on time. With Nygård's ever-expanding use of information technology, tracking product performance across store locations and its reseller channel is, of course, an absolute necessity in enabling greater efficiency at an organization that runs around the clock.

Selecting MicroStrategy Mobile

Nygård has been relying on MicroStrategy since 2000 for all of its business intelligence initiatives. When the company decided to build out its business intelligence platform to support a mobile workforce, selecting MicroStrategy Mobile was a clear-cut decision. Nygård wanted to empower associates in the field by providing the insight and information on their BlackBerry devices so they could make the right operational decisions wherever they were. This would free them from having to sit at a desk in front of a 17-inch LCD screen attached to a desktop PC.

The application delivers on what Peter Nygård himself believes: that all desktop applications need to extend to a mobile world. With MicroStrategy Mobile delivering the right information to BlackBerry users in stores and to associates on the road, selling has never been easier.

Building the Solution

Nygård started its own mobile application in September 2006 by building a prototype of its retail fashion intelligence application. This mobile product lookup application was built in about eight hours and demonstrated the vision of what the company wanted. After showing it to MicroStrategy executives, Nygård product teams started exchanging ideas and lessons learned. MicroStrategy was helpful in guiding the teams through their development process, and business users, in turn, provided valuable feedback and drove enhancement changes to help MicroStrategy improve their product.

Once MicroStrategy Mobile was released in September 2007, the implementation was simple. Nygård already was sending many reports via MicroStrategy Narrowcast Server as high-level summaries in e-mails, with additional Excel workbook attachments that included supporting data. The challenge with these e-mail reports was that users were tethered to their computers, having to wait for reports to become available and then print select reports, before heading out to the field.

Ease of use and one-click access to real-time information headed the list of what the company's mobile users wanted. MicroStrategy Mobile delivered on these requirements. It also

Figure 1. With MicroStrategy Mobile delivering the right information to BlackBerry users in the field, selling has never been easier.

replicated and delivered to Nygård's BlackBerry devices the same reports that users had been receiving via e-mail. The reports retained a consistent look and feel—for example, consistent headings and standards for KPI abbreviations, the same colors that define what users see, and the same special fonts to highlight totals and subtotals.

Immediate ROI

Fortunately, the migration from the e-mail–based system to MicroStrategy Mobile was fast and seamless because Nygård did not need to change any reports. The ROI was immediate; since the reports were always available, the company gained productive time. No more waiting!

MicroStrategy Mobile was able to bridge the gap and reach Nygård's mobile workforce. Today, the company's 140 BlackBerry users have a wide range of interactivity—just as they would at their desks—and are asking for more reports. Nygård's business needs have been met, and the company is thrilled with this remarkable product and its success across its user community. ●

For a free white paper on this topic, download "Reducing Total Cost of Ownership: Delivering Cost Effective Enterprise Business Intelligence," or click here for more information about MicroStrategy.

LESSON FROM THE EXPERTS

Developing BI Applications in a Heterogeneous Data Environment

By Bala Chandran
Senior Product Manager, MicroStrategy

Most organizations initially built business intelligence (BI) applications as small departmental solutions targeting specific business processes. Applications such as product sales reporting and HR compensation reporting were used by a limited number of power users to access a single, well-defined data source. Organizations now realize the transformative power of BI and seek to establish a single, enterprisewide reporting and analytics architecture that enables all users to make better-informed business decisions.

As architects continue to design these enterprise BI applications, they find that critical corporate information is stored in hundreds of different databases, flat files, cubes, and other data stores. In order to make effective, timely decisions, users need access to data from these multiple sources to provide an integrated view of business performance. It is important that users can view high-level dashboards, drill to detailed data, and receive proactive alerts—all without worrying about the source of the data. Organizations should look closely at how BI platforms can support heterogeneous data source environments before making their technology selections.

Enterprise Data Warehouses vs. Federated Data Stores

Creating a single enterprise data warehouse (EDW) has traditionally been the optimal solution to consolidating enterprise information. However, organizations have found that new data sources appear in the IT landscape faster than can be incorporated into the EDW. In contrast, some organizations have opted for a distributed architecture with information stored in several data marts and joined at the query level, often referred to as federated data. While both approaches have pros and cons, in either case BI software must effectively access multiple data sources to satisfy increasingly complex user requirements.

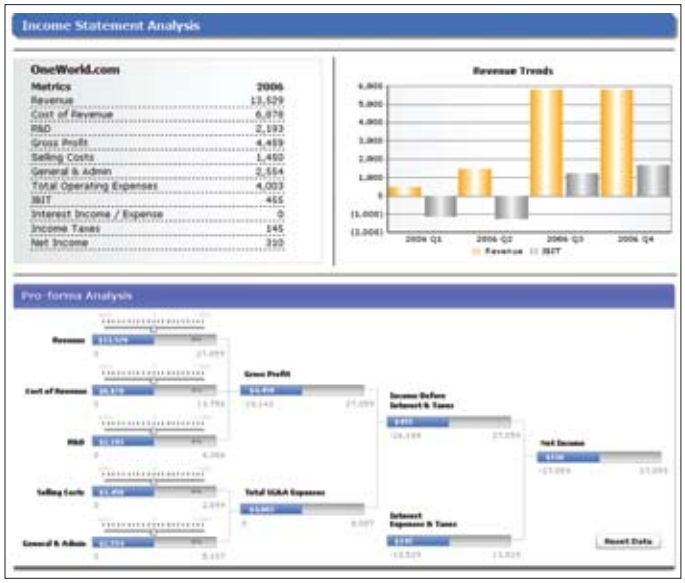


Figure 1. Dashboards often require access to as many as 10 different data sources in order to present a holistic business performance perspective.

Dashboards

A true enterprise BI platform would be incomplete without the ability to access and query multiple information sources and present the information through a single standard interface. The user should be able to navigate across all five styles of BI—scorecards and dashboards, enterprise reports, OLAP analytics, predictive analytics, and alerting—irrespective of the data source. Then, business users can concentrate on solving business problems, rather than understanding data sources.

Dashboards, in particular, are an ideal way to monitor corporate performance across a heterogeneous data environment. The highly interactive and visual nature of dashboards, such as MicroStrategy's Dynamic Enterprise Dashboards™, appeals to businesspeople who can use them to identify problem areas quickly and drill for further information. From a development standpoint, dashboards often require access to as many as 10 different data sources in order to present a holistic business performance perspective.

Dashboards must be interactive and highly visual to help users understand business performance. Dashboards that access multiple data sources typically fall into two categories. In the first case, a single dashboard can present data from many sources without joining data across those disparate sources. Overall enterprise KPIs can be monitored using this type of dashboard. In the second case, a single dashboard must join the data across heterogeneous data sources to perform the required analysis. For example, actual sales and forecast sales may be in two different systems. In this case, the BI layer must join the data across from the appropriate data sources. Of course, to perform joins between multiple data sources, common keys need to exist between heterogeneous data sources.

Summary

As dashboard-based performance management solutions become more prevalent throughout the enterprise, architects and developers should consider how they could integrate data from multiple sources into BI applications. BI platforms, like MicroStrategy, provide a compelling solution to integrate disparate data sources through highly interactive, dynamic dashboards. ●

LESSON FROM THE EXPERTS

Introducing a Data Warehouse for Event Data

HOW A COLUMNAR-BASED DATA WAREHOUSE OFFERS SUPERIOR PRICE AND PERFORMANCE OPTIONS FOR EVENT DATA

By Ed Chopskie
Vice President of Marketing, SenSage, Inc.

Event data, once commonly referred to as "audit trails," is data or a set of records sequenced chronologically. Event data contains evidence directly pertaining to and resulting from the execution of a business process or system function. Common event data captured and retained by organizations includes, for example, records resulting from activities such as business transactions or communications by individuals, systems, accounts, or other entities. Business transaction examples include event records created from banking transactions, updates to shipping status, historical prices, and radio frequency identification records. Examples of communication records include call detail records of telephony and internet traffic/transaction data by governments and commercial organizations.

Long-term retention of these types of records is often required to detect fraud or to analyze performance trends. In the case of communication records, the retention is mandated by government regulations for access by law enforcement agencies. One such communication regulation is Directive 2006/24/EC, which was passed by the member countries of the European Union in March 2006 and requires the member states to ensure that communications providers must retain, for a period of time between six months and two years, necessary data as specified in the directive.

Regardless of the source of event data, the data shares common characteristics. Event data is voluminous, and current implementations have broken through hundreds of terabytes and are approaching a petabyte. Event data is written once and never updated, as audit trails must never be modified; as a result, the use of data warehouse solutions built on relational database technologies originally designed for supporting OLTP is extremely efficient. Additionally, event data is always inserted and later searched on the basis of time, introducing storage and querying challenges that most relational databases do not easily support. Finally, event data is typically "flat" with many distinct columns and not subject to normalization.

Because of the volume and nature of the data being stored, traditional database warehouses quickly become impractical to use for event data for a number of reasons, including performance and cost.

The purpose of this lesson is to introduce a new type of data warehouse, an event data warehouse (EDW), built on a patented columnar database, to provide superior performance for processing event data—at a cost that is an order of magnitude less than traditional solutions.

Introduction to Columnar Databases

Before describing how the SenSage EDW provides a complete solution for collecting, storing, and analyzing event data, a brief description of columnar databases is required. This introduction will help differentiate the architectural differences between columnar databases and relational databases for the purpose of illustrating how a columnar organization is superior in performance and price for event data; it is not meant to be exhaustive.

A columnar database organizes data by columnar, rather than row, format used by relational database management systems. While the difference may sound trivial, a columnar architecture provides distinct advantages for certain classes of data, including event data. Data for each column is stored together, and this provides performance gains by allowing queries to reference only the data selected. Indexes are unnecessary in columnar databases, as each column is actually an index, thus significantly reducing the storage and maintenance requirements of relational databases. Additionally, data in columns provides a massive opportunity for data compression as the data in columns is similar, unlike the compression of an entire row with different data types.

Columnar databases are not new; several have existed on the market for some time in relative obscurity compared to relational databases.

The SenSage EDW Solution

While SenSage is built on a patented[1] columnar database, the entire solution provides all components required for event data warehousing, including the ETL and analytics layer. Additionally, the SenSage EDW supports open access from a number of methods including SQL, Perl DBI, and JDBC.

While the SenSage EDW is practically a turnkey solution for event data requiring very little knowledge of columnar database administration, the underlying architecture is extremely advanced, utilizing a distributed, clustered, share-nothing architecture built from the ground up for event data.

The SenSage EDW is a software solution for deployment on commodity server "nodes." This deployment model allows the SenSage

1. U.S. Patent #7,024,414 can be referenced at www.uspto.gov/patft

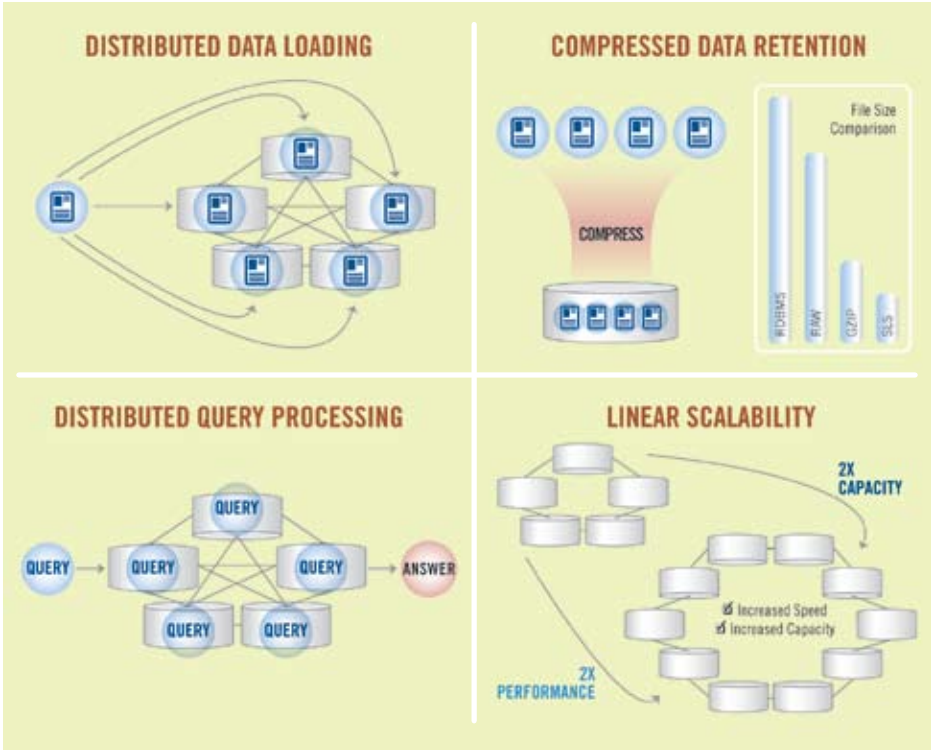


Figure 1. Components included in the SenSage EDW architecture

ETL data loader to spread data across nodes allowing for record insertion rates of hundreds of thousands of records per second. As data is loaded, it is compressed at a typical rate of 10:1. Data queries are also distributed across the data warehouse nodes, providing increased performance. As columnar database, the SenSage EDW does not require indices but provides an advanced querying technique, known as Bloom Filters, to determine if the data required to satisfy a query exists in trees on the nodes. Using this combination of advanced features, the solution has been deployed and proven to be capable of querying over 30 million records a second, yielding query results in minutes. This performance is typically an order of magnitude faster than relational database management system capabilities.

Nodes can easily be added to the EDW cluster as data storage and query requirements increase. Additionally, since the SenSage EDW is delivered as software and not as a proprietary appliance, organizations can take advantage of performance increases in hardware and storage as they become available. The SenSage EDW supports storage of event data locally on the nodes themselves or via SAN/NAS solutions.

Conclusion

While traditional data warehouse solutions based on relational databases are very often the most practical solution for most business requirements, they are ill equipped to support the unique data characteristics of event data. Utilizing traditional solutions for event data is possible but at a price and performance differential that cannot be justified. Columnar databases are well suited to processing event data, and there are many columnar database management systems on the market, including open source solutions such as C-Store.

The SenSage EDW takes advantage of a columnar architecture but expands the solution by providing an architecture that provides a turnkey approach to managing event data:

- It includes the ETL applications within the product. Administrators can quickly map their event data sources without knowledge of the underlying table format or database administration skills. There is no requirement to obtain third-party ETL tools or expertise.

- SenSage provides analytics and query-building wizards that do not require users to have any knowledge of SQL or any other programming languages. Other access methods to the data are supported, but most users do not find the need to implement third-party query tools.

- It is deployed as software and takes advantage of implementations on commodity clustered server solutions. Customers are not locked into proprietary appliance-based solutions and therefore can easily increase performance by utilizing new technologies as they become available.

SenSage has been sucessfully deployed at more than 300 sites. With well over a petabyte under management, the SenSage EDW commonly supports implementation of more than 100 terabytes, with some scheduled to reach a petabyte at current growth rates. Unlike emerging or open source solutions, the SenSage EDW is a proven and supported solution used worldwide by Global 2000 organizations.

The SenSage EDW is a powerful and economic solution for event data when compared to traditional data warehouses.

For a free white paper on this topic, download "Event Data Warehousing," or click here for more information about SenSage.

CASE STUDY

Enhanced Business Intelligence—On Demand

Commentary by Ken Rudin
CEO, LucidEra

Business intelligence (BI) holds the promise of delivering competitive advantage by giving managers insight into what customers want and how the business is performing. But the value of business intelligence can be hard to realize. First, in today's interconnected world, application and process outsourcing means that critical information may exist outside an organization's walls. And many companies lack the resources—time, skills, or budget—to pull together data coming from different sources and different formats.

LucidEra was founded in 2005 to offer a new approach to BI: an on-demand service that delivers prebuilt analytic applications that are simple to set up, simple to use, and simple to buy. LucidEra's customers are companies that want better insight into their sales processes. LucidEra's business analytics-as-a-service model combines information from a company's CRM and financial applications and then delivers that information in dynamic reports, on demand.

One of the challenges to LucidEra's business model was integrating information from multiple applications and databases. For any customer-related analytics that combine data from multiple application sources, the company's software-as-a-service (SaaS) platform needed to know how to identify a customer uniquely across both systems. For example, if the company brings in opportunities for the company IBM from salesforce.com, and orders for International Business Machines from Oracle E-Business Suite, they need to know that those two names represent the same company.

Duplicate records for the same customer have always been a challenge to CRM adoption and can have an even more profoundly negative effect on information analysis and business intelligence success. For example, one sales rep may create an entry for IBM and associate opportunities to IBM to track the sales process. Another sales rep may create an entry for I.B.M. and do the same. They can meet their individual needs by creating two separate names, but it complicates larger analysis and can lead to incorrect business decisions.

Technology from Identity Systems helps LucidEra solve this problem and deliver a completely hosted analytics service to its customers. Once customer data is loaded into LucidEra's SaaS platform, they use Identity Systems' matching algorithm to identify duplicate sales rep and account records. This improves the quality of the data, which results in trusted information and analysis for subscribers.

From Many, One

The Identity Systems solution helps LucidEra focus on better sales and customer analysis without having to worry about the nuances of the technology. The Identity Systems solution met all the requirements that LucidEra had set out. It provided proven and extensible technology. The software integrated with a variety of systems and developer resources were widely available. Finally, there was scalability to provide future growth and meet LucidEra's performance requirements.

LucidEra implemented the Identity Systems solution in 2006. The software integrated easily and quickly with existing systems and permitted the company to offer the industry's first complete on-demand business analytics service. The Identity Systems solution has helped to automate the challenge on name matching—something you simply can't substitute with manual labor.

The new approach gives LucidEra an edge in the marketplace. The built-in account name matching—along with the SaaS analytical capabilities—makes a compelling offering that helps to differentiate it from other solutions. But in the end, perhaps the greatest benefit goes to LucidEra customers. LucidEra is providing them with accurate information and analyses that they can rely on and communicate throughout the business. Business needs a 360-degree view of customers, and Identity Systems technology is a critical enabler in LucidEra's solution to help them meet that need. ●

For a free white paper on this topic, download "Data Warehousing, Next-Generation Business Intelligence, and the Evolution of Data Quality," or click here for more information about Identity Systems.

LESSON FROM THE EXPERTS

New Frontiers in Identity Resolution

By Jim Jarvie
Marketing Director, Identity Systems, Inc.

Identity resolution—the process of determining which data representations refer to the same entity—traditionally focused on matching name and address records from company systems. But today's businesses face a new set of challenges:

- Matching against external lists. These may be watch lists from government authorities, fraud lists from industry consortia, shared customer lists from business partners, or data acquired through mergers and acquisitions. All are produced by organizations whose formats, standards, and processes are beyond your control, so the traditional strategy of improving data quality at the source is not an option.

- Unstructured sources. Information may be gathered from Internet search strings, field reports, telephone transcripts, e-mail and text messages, Web pages, and other unstructured formats. Parsing these to extract useful information and dealing effectively with the inevitable ambiguities is a major new challenge.

- Multiple geographies. Companies must increasingly integrate data from different countries, languages, and even character sets. Each region has its own formats, rules, and local information. Data from different regions is often mingled within a single input file, forcing the identity resolution system to determine on a record-by-record basis which set of rules should apply.

- More types of information. Individuals may be matched using an e-mail address, Internet cookies, Web sites, device IDs, product serial numbers, GPS coordinates, and other identifiers beyond the traditional name and postal address. And it's not just individuals: companies increasingly use identity resolution to track products, materials, vehicles, equipment, and even legal documents.

- More applications. Identity resolution results are now served back for purposes ranging from compliance to marketing to customer service. Each application has its own ideal balance between cost, speed, and accuracy. Privacy and regulatory rules often mean that the data available for different purposes will vary as well.

- Need for quick response. Today's systems increasingly must react—in real time—to input from a telephone agent, Web site, kiosk, or retail associate. Adding to the challenge, the data from these sources is often user-entered, meaning it is less accurate and consistent than entries from trained company employees.

No one technique can meet all of these new requirements. Some approaches that help include:

- Using multiple keys. In a perfect world, a single match key could combine different data elements in a fixed sequence to associate related records. But missing data, ambiguous meanings, alternative identifiers, and inconsistent formats make this impossible. Multiple keys ensure that related information is spotted even when it is hidden in different locations on the input records.

- Selectable search levels. Because different applications have different requirements for the accuracy, cost, and response time, the identity resolution system must make it easy to change the balance among these on the fly.

- Extensive local reference sets. Sophisticated string matching by itself can never substitute for standardization based on local rules and reference data. This becomes increasingly critical as each system handles a broader range of geographies and data types.

- Efficient adjustment to new needs. New data types, formats, geographies, and applications are added at an ever-increasing rate. The identity resolution system must accommodate these quickly and effectively. This implies functions for importing and evaluating the new data, training the system to use it, testing the results for accuracy and assurance that they don't cause problems elsewhere, and simple mechanisms for making the results accessible.

Not every system can meet these new requirements. But the cost of acquiring suitable technology is the price of entry to a world where identity resolution adds new value to enterprise systems—providing benefits that vastly exceed the cost of the ticket itself. ●

CASE STUDY

Booking a Better Customer Experience

By Tina Wefer
Senior Manager, Product Marketing, Initiate Systems

Overview

One of the nation's top bookseller brands sells more than 300 million books a year through eight million orders online and in-store. On a daily basis, it handles an average of 100,000 customer service requests. Customer data is housed in five core customer systems: special orders, educators, institutions, members, and online business.

Challenge

As part of a CEO directive to improve the customer experience, the company wanted to streamline and enhance its customer interactions. To facilitate this, it needed a complete customer view across more than 700 stores, its member program, and the Web site.

Previously, when a customer contacted a store, there was no visibility into his membership status or orders from other stores. Every time an order was placed, he was asked for the same contact information. This frustrated customers and often created duplicate records. If a customer gave a different e-mail address, the problem compounded, spurring conflicting or duplicate e-mails to the customer. For example, if a customer opted out of receiving special offers via e-mail with one e-mail address, the other address might continue to receive the offers.

By enabling a complete view of customer data scattered across these sources, the bookseller had a number of goals:

- Avoid duplicate and inconsistent e-mails while reconciling customer opt-out preferences and improving privacy compliance
- Enable real-time customer recognition by leveraging full customer views, helping associates at 700+ stores better serve customers by viewing order history and status
- See relationships between customers, such as those living in the same household
- Implement this customer hub without affecting existing operations

Solution

Improving the customer experience requires first recognizing and understanding customers and behavior, then leveraging that information for marketing, customer retention, and other initiatives. With a master data management (MDM) solution that matches and links records across databases, a customer's records can be united by common demographic information. For example, William Scott's records can be matched with those of Bill Scott when they share a common e-mail address or postal code—regardless of whether Mr. Scott has his member card in hand when he places an order.

With Initiate software in place, the company will achieve a litany of initiatives, including:

- Supporting customer recognition, more effective e-mail communication, and privacy preference management by better identifying unique customers
- Improving the customer experience at all points of service
- Targeting marketing efforts to recognize and reward loyal customers
- Understanding customers, their records, and their history through focused customer search capabilities across the enterprise
- Improving the accuracy of business intelligence reporting and customer segmentation by correctly linking customer history to unique customers
- Laying a foundation for future success and additional customer data sources

Results

First, an Initiate JumpStart evaluated approximately 40 million customer records from three sources, identifying approximately 11 million duplicates. Each of these duplicates indicated an opportunity for a customer's records to be linked. Having a single view greatly increases marketing abilities and offers chances to build the customer relationship. Rather than asking for the same demographic information, the bookseller can offer related items or programs.

Next, the bookseller implemented Initiate software as its enterprisewide MDM solution. With data hubs to manage both customers and households, it immediately saw real results. As associates at individual stores can see full customer views, including complete order history across the company, customer service is improving.

Moving forward, this bookseller plans to leverage its Initiate solution to add additional data sources and further streamline operations, while continuing to improve customer service. With Initiate software, this company is poised to remain on top of the bookselling heap. ●

For a free white paper on this topic, download "Master Data Management and Accurate Data Matching," or click here for more information about Initiate Systems.

LESSON FROM THE EXPERTS

Getting Started with Master Data Management

By Tina Wefer
Senior Manager, Product Marketing, Initiate Systems

Think of it as a highly accurate tool for making decisions and gaining insight about customers, products, employees, vendors, and other aspects of a business with laser-like precision. Master data management (MDM) systems generate and maintain an enterprisewide "system of record" that contains consistent, reliable information necessary to perform vital business functions. MDM has become an increasingly hot topic due to its ability to increase revenue, reduce costs, improve customer service, and make it easier to comply with regulations.

Though experts agree that the holy grail of MDM is a solution that unites data from diverse applications (e.g., sales, support, billing, marketing), lines of business, and channels, developing such an advanced IT architecture typically takes years, is extremely expensive, and virtually eliminates the chance for a fast return on investment (ROI). Some companies approach MDM as a data warehousing or system consolidation project, assuming they need to merge all data before taking advantage of it. A more reasonable approach is to implement a phased MDM strategy that enables enterprises to develop an effective working model for future development, while allowing a fast ROI on the first phase of the project. Consider starting with a registry or hybrid implementation; by leaving the data where it currently resides rather than consolidating it into one master source, you can achieve quicker implementations and quicker time to value. You can evolve to a transactional implementation over time if your needs dictate.

Knowing Where to Start

While businesses can choose from a number of data domains on which to focus, customer data is a natural starting point for an incremental MDM approach for many companies.

Figure 1. Master data management systems generate and maintain an enterprisewide "system of record" that contains consistent, reliable information necessary to perform vital business functions.

According to a 2006 research study by The Data Warehousing Institute, information about customers is the data most requested by users.

An incremental customer-centric approach to MDM can significantly enhance customer service efforts by providing more accurate and complete customer data and reducing the risk of human error. Access to accurate, real-time master customer data gives sales, marketing, and service teams a better view of customers and their preferences, enabling them to provide targeted offerings and personalized service that improve customer relations.

Selling to Management

When selling a plan of attack to executives, remember they need to see a return on investment in a relatively short time frame. Make it easy for executives to envision the value and understand long-term goals while focusing on short-term gains. Following are some tips:

- Sketch the big picture, but concentrate on short-term gain
- Identify three projects that overcome inertia and demonstrate tangible business value
- Build detailed business cases with projects that people can rally around
- Demonstrate an understanding of effects on major business processes
- Evaluate projects from both an IT and business case point of view
- Demonstrate short-term gains to sell long-term opportunities
- Commit to realistic dates as you plan

Final Thoughts

The power of MDM derives from its ability to manage data across a company's entire information infrastructure. However, many companies lack the means or resolve to implement MDM at once across all applications, data sources, and physical locations in their enterprise. For these organizations, a phased approach is more cost effective. And for a good many companies, the first launching point for MDM should be the customer realm. Businesses that successfully implement a customer-centric MDM solution in one area of the company can then move on to other areas as business requirements demand, giving them the intelligence they need to improve business processes, build revenues, and increase competitiveness in the years ahead. ●

LESSON FROM THE EXPERTS

Getting Off on the Right Foot: Avoiding Common Master Data Management False Starts

By Ravi Shankar
Director, Product Marketing, Siperian, Inc.

Companies wishing to start a master data management (MDM) project may be unsure where and how to begin. After all, MDM is a journey, and success or failure at the first step either defines or dooms the further evolution of the project. Recently, industry analysts have been recommending a cautious approach to starting with MDM—suggesting that companies start with a single data type (such as customer), implement MDM using a small footprint (such as registry style), or deploy MDM solely with a data warehouse to improve reporting. Inherently, these technology-focused approaches reduce project risk and relieve the data governance burden. Companies may readily adopt these approaches as perfectly reasonable starting points and lean to a more risk-averse approach to their initial MDM implementation in hopes of mitigating risks. However, these same approaches may limit the scope and potential return on investment (ROI) from MDM, since they do not attempt to solve the most pressing and difficult business problems.

Beware of Technology-Focused Starts

A nearsighted focus only on the technology aspects of MDM may ultimately lead to minimal business adoption and therefore may severely constrain the business ROI. The following business case scenarios illustrate how three different technology-focused approaches will limit MDM's usefulness in solving difficult business problems.

Restricting MDM to a single master data type. For example, an MDM solution that is deployed to solve buy and sell-side supply chain processes and more effectively manage the procurement of direct and indirect materials and the distribution of products, necessarily needs to involve managing vendor, customer, material, and product master data. Starting with only one of these master data types will not effectively improve the systemic supply chain and would severely constrain the usefulness of an MDM solution for supply chain performance management.

Confining MDM to a registry approach. An MDM solution implemented to improve credit risk management and capital requirements for compliance with Basel II regulations will need to reconcile conflicting counterparty master data and legal hierarchies and store them centrally for immediate access. In this case, a registry approach would only identify counterparties as duplicates without determining a system of record or the correct definition for the counterparty. As a result, credit risk managers would be unable to determine which counterparty definition is current and accurate. In addition, a registry approach cannot determine the legal hierarchies required to calculate the aggregated risk exposure. Consequently, the credit risk managers would need to go through a process to determine the correct entry and sometimes combine information from different systems to arrive at a single definition and legal hierarchy representation. From that point forward, the information would act as the single best source of information for all credit risk calculations. A small footprint using the registry approach would not effectively solve this difficult business problem.

Limiting MDM to analytical usage. In the case of using MDM to improve order-to-cash, reliable master data needs to be synchronized back to operational systems, such as order management, in order to enhance the business process. Where the master data is only synchronized to a data warehouse, the efficacy of the order-to-cash business process cannot be improved, since this process is inherently operational in nature. Measurable hard dollar benefits derived from MDM are only achievable with business process improvements.

Taking a technology-focused approach may enable your organization to get started with MDM quickly, but it may not effectively solve the difficult business problems or deliver the requisite business value. In fact, the resulting solution more readily runs the risk of being perceived by business users as yet another IT initiative unable to address their business needs. This will make it increasingly difficult to further evolve or extend the solution—boding a premature death for the enterprise MDM initiative or, even worse, "getting stuck at the gate."

It is important also to take notice that some MDM vendor solutions support only a single architecture style, such as registry, or can be deployed only for a single usage—either operational or analytical. These solutions simply cannot be extended to other architectural styles or another usage mode, which can severely limit their usefulness in addressing the most challenging of business problems. In addition, a technology-centric start will not fulfill the most important needs around enterprise master data governance.

Master your data. Master your business.

Start with the Business in Mind

MDM is more precisely about solving business problems by efficiently managing master data that is critical to a company's business operations. Consequently, how an MDM solution is implemented depends foremost on which business problems are being tackled. Only a business-focused approach can provide a complete MDM solution that addresses the specific business problem and provides tangible business value and significant ROI in a short-term timeframe. By taking this approach, you can ensure the success of your MDM initiative and pave the way for expansion across the organization. How to get started? A pragmatic place to begin is to answer these three questions:

1. Which business problems need to be tackled? Organizations should start by first identifying the business processes that are inefficient, and among those, which ones should be addressed first. By choosing a business process to start with, the master data types that need to be managed will become evident. For example, two business processes within a company's supply chain are experiencing problems—different divisions within the company are procuring direct materials from the same vendor at different contracted rates, and sales people are competing for the same customer's business. The master data integral to improving these business processes are vendor, contract, customer, materials, and product information.

2. What is the business use? Next, identify how business users will use the master data within their business processes in order to determine the most appropriate architectural style and usage modes to support the needs of the business users. As an example, in order to ensure that the same contracted rates for procuring different direct materials from a supplier are made available at different touch points, the MDM system needs to reconcile conflicting vendor, contract, and direct materials data and then centrally store it. (Analysts refer to this architectural style as coexistence or transactional.) The data also needs to be made available to the supply chain and contract management systems. And to ensure sales alignment, the MDM system needs to make customer and product information available to the data warehousing system for accurate and timely analysis and reporting.

3. What are the business requirements for master data governance? Finally, it is important to understand the business requirements for governing the master data in order to determine the requirements for master data availability, usability, integrity, and security. For instance, the procurement department will require a high degree of integrity for vendor and contract data and will need to be able to make this data available to procurement agents in real time. The contract negotiation team, on the other hand, may require the same degree of data integrity, yet not in real time. Similarly, the sales team would have a requirement that only sales managers are able to perform sales force alignment, while sales representatives only have access to information for their assigned territory.

The Right Start Ensures an Initial MDM Win

What becomes obvious from these and other examples you may consider in your business is that MDM will almost always require a multi-entity deployment (such as customer and product) and an architectural style that is not restricted to registry alone. In most instances, synchronization with both operational and analytical systems may also be essential to effectively address the specific business needs of your organization.

By taking a business-focused approach to MDM, you can provide a complete solution to the most challenging of business problems—using only the required master data, implemented with the correct solution architecture, deployed for the correct business use, and with the correct data governance structure. When a pressing business problem is successfully solved by an MDM solution, adoption of the solution dramatically increases among business users because it eliminates inefficiencies and improves productivity—resulting in measurable cost savings and higher ROI. Starting with a defined business problem allows you to start small so that success can be demonstrated before expanding the solution to other business units, geographies, or divisions. Once business users experience the benefits of an MDM solution, they will more readily support its use in other areas—paving the way for an enterprise MDM solution. ●

For a free white paper on this topic, download "How to Write an RFP for Master Data Management: Ten Common Mistakes to Avoid," or click here for more information about Siperian.

Strategies for Managing Spreadmarts

Migrating to a Managed BI Environment

BY WAYNE W. ECKERSON AND RICHARD P. SHERMAN

Business users are empowered by knowledge—and knowledge comes, in part, from having access to accurate and timely information. It is generally up to the information technology (IT) department to supply this information. But it doesn't always work out that way.

Definition of a Spreadmart. TDWI used the following definition of a spreadmart in the survey it conducted as part of this report:

> A *spreadmart* is a reporting or analysis system running on a desktop database (e.g., spreadsheet, Access database, or dashboard) that is created and maintained by an individual or group that performs all the tasks normally done by a data mart or data warehouse, such as extracting, transforming, and formatting data as well as defining metrics, submitting queries, and formatting and publishing reports to others. Also known as *data shadow systems*, *human data warehouses*, or *IT shadow systems*.

In organizations all over the world, business people bypass their IT groups to get data from spreadmarts. Spreadmarts are data shadow systems in which individuals collect and massage data on an ongoing basis to support their information requirements or those of their immediate workgroup. These shadow systems, which are usually built on spreadsheets, exist outside of approved, IT-managed corporate data repositories, such as data warehouses, data marts, or ERP systems, and contain data and logic that often conflict with corporate data. Once created, these systems spread throughout an organization like pernicious vines, strangling any chance for information consistency and reliability. You'll find them in all industries, supporting all business functions. According to TDWI Research, more than 90% of all organizations have spreadmarts. (See Figure 1.)

DOES YOUR GROUP HAVE ANY SPREADMARTS?

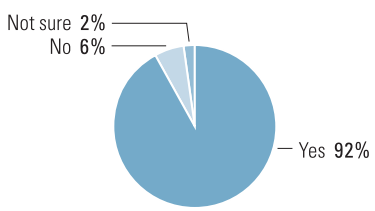


Figure 1. Status of spreadmarts

Spreadmarts often lead to the phenomenon of dueling spreadsheets. Murray Trim, a management accountant with Foodstuffs South Island Limited, described one such situation: "We have had the classic situation of two people presenting ostensibly the same data at a board meeting with different figures, which they got from different spreadmarts." Donna Welch, a BI consultant at financial holding company BB&T, talks about the issues of trust that arise from dueling spreadsheets: "We constantly hear our users talk about management's distrust of their reports because multiple people came up with different answers."

Who and Why. Spreadmarts are usually created by business analysts and power users who have been tasked to create custom reports, analyses, plans, benchmarks, budgets, or forecasts. Often, these analysts—especially those in the finance department and the financial services industry—have become proficient with Microsoft Excel or Microsoft Access and prefer to use those tools to create reports and analyses. As a result, most are reluctant to adopt a new corporate reporting "standard," which they believe will limit their effectiveness. Change comes hard, especially when it means learning a new toolset and adapting to new definitions for key entities, calculations, or metrics. Executives perpetuate the problem because they don't want to pay hundreds of thousands of dollars or more to build a robust data infrastructure and deploy enterprise reporting and analysis tools. Instead, spreadmarts proliferate.

Dangers of Spreadmarts

Inconsistent Views. The problem with spreadmarts is that their creators use different data sources, calculations, calendars, data conversions, naming conventions, and filters to generate reports and analyses based on their view of the business. The marketing department views customers and sales one way, while the finance department views them another way. The way the business operates in Germany is different from the way it operates in Brazil. Business units sell the same products with different names, packaging, pricing, and partner channels. When each group manages its own data and processes, it's nearly impossible to deliver a consistent, enterprise view of customers, products, sales, profits, and so on. These parochial silos of data undermine cross-departmental and business unit synergies and economies of scale.

Excessive Time. In addition, business analysts spend two days a week—or almost half their time—creating spreadmarts, costing organizations $780,000 a year! Instead of analyzing data, these high-priced employees act like surrogate information systems professionals, gathering, massaging, and integrating data. Many executives have initiated BI projects simply to offload these time-consuming data management tasks from analysts.

Increased Risk. In addition, spreadmarts are precarious information systems. Because they are created by business users, not information management professionals, they often lack systems rigor. The problems are numerous:

- Users often enter data into spreadmarts by hand, which leads to errors that often go undetected.
- Few spreadmarts scale beyond a small workgroup.
- Users may create poorly constructed queries, resulting in incorrect data.
- Spreadmarts may generate system and data errors when they are linked to upstream systems or files that change without notice.
- Users embed logic in complex macros and hidden worksheets that few people understand but nevertheless copy when creating new applications, potentially leading to unreliable data.
- There is no audit trail that tracks who changed what data or when to ensure adequate control and compliance.

In short, spreadmarts expose organizations to significant risk. Business people may make decisions based on faulty data, establish plans using assumptions based on incorrect analyses, and increase the possibility of fraud and theft of key corporate data assets.

Not All Bad?

No Alternative. Despite these problems, there is often no acceptable alternative to spreadmarts. For example, the data that people need to do their jobs might not exist in a data warehouse or data mart, so individuals need to source, enter, and combine the data themselves to get the information. The organization's BI tools may not support the types of complex analysis, forecasting, or modeling that business analysts need to perform, or they may not display data in the format that executives desire. Some organizations may not have an IT staff or a data management infrastructure, which leaves users to fend entirely for themselves with whatever tools are available.

As such, spreadmarts often fill a business requirement for information that IT cannot support in a timely, cost-effective manner. Spreadmarts give business people a short-term fix for information that they need to close a deal, develop a new plan, monitor a key process, manage a budget, fulfill a customer requirement, and so on. Ultimately, spreadmarts are a palpable instantiation of a business requirement. IT needs to embrace what the business is communicating in practice, if not in words, and take the appropriate action. Thus, spreadmarts should not be an entirely pejorative term.

Cheap, Quick, Easy. Moreover, since spreadmarts are based on readily available desktop tools, they are cheap and quick to build. Within a day or two, a savvy business analyst can prototype, if not complete, an application that is 100% tailored to the task at hand. Although the spreadmart may not be pretty or "permitted," it does the job. And it may be better than the alternative—waiting weeks or months for IT to develop an application that often doesn't quite meet the need and that costs more than executives or managers want to pay.

Nevertheless, there is a high price to pay for these benefits in the long term. Many executives have recognized the dangers of spreadmarts and made significant investments to fix this problem. However, not all have succeeded. In fact, most struggle to deliver a robust data delivery environment that weans users and groups off spreadmarts and delivers a single version of truth.

Remedies

Managed BI Environment. The problem with spreadmarts is not the technology used to create them. Spreadsheets and other desktop-oriented tools are an important part of any organization's technology portfolio. The problem arises when individuals use these tools as data management systems to collect, transform, and house corporate data for decision making, planning and process integration, and monitoring. When this happens, spreadmarts proliferate, undermining data consistency and heightening risk.

The technical remedy for spreadmarts is to manage and store data and logic centrally in a uniform, consistent fashion and then let individuals access this data using their tools of choice. In other words, the presentation layer should be separated from the logic and data. When this is done, business users can still access and manipulate data for reporting and analysis purposes, but they do not create new data or logic for enterprise consumption. At TDWI, we call this a managed business intelligence environment. The goal is to transform spreadmarts into managed spreadsheets. This lets IT do what it does best—collect, integrate, and validate data and rules—and lets business analysts do what they do best—analyze data, identify trends, create plans, and recommend decisions.

BI vendors are starting to offer more robust integration between their platforms and Microsoft Office tools. Today, the best integration occurs between Excel and OLAP databases, where users get all the benefits of Excel without compromising data integrity or consistency, since data and logic are stored centrally. But more needs to be done.

Change Management. Applying the right mix of technology to address the spreadmart problem is the easy part. The hard part is changing habits, perceptions, behaviors, processes, and systems. People don't change on their own, especially when they've been successful with a certain set of tools and processes for analyzing data and making decisions. Changing a spreadmart-dependent culture usually requires top executives to both communicate the importance of having unified, consistent, enterprise data, and to apply incentives and penalties to drive the right behaviors. Ultimately, change takes time, sometimes a generation or two, but the right organizational levers can speed up the process.

Aligning Business and IT. Another dynamic driving spreadmarts is the lack of communication and trust between business and IT. The business doesn't adhere to the architectural standards and processes designed to support its long-term interests, while IT doesn't move fast enough to meet business needs. To reverse this dynamic, both business and IT must recognize each other's strengths and weaknesses and learn to work together for the common good. IT must learn to develop agile information systems that adapt quickly to changing business conditions and requirements. The business must recognize the importance of building sustainable, scalable solutions. IT must learn about the business and speak its language, while the business must not blame IT for failures when it continually underfunds, overrides, and hamstrings IT so that it cannot possibly serve business needs.

Architectural Approaches to Spreadmarts

Recognizing that you have a spreadmart problem is the first step. Most of the people we surveyed know their organizations have spreadmarts, but they don't know what to do about them.

The survey presented respondents with nine different approaches to addressing the spreadmart issue. (See Table 1.)

Ironically, the most common approach that organizations use is simply to leave the spreadmarts alone. But as with everything else in life, ignoring a problem does not make it go away, and often makes it worse. When asked how effective this approach was, a majority (58%) said "not very effective."

Replace with BI Tools. The next most popular approach is to "provide a more robust BI/DW solution," employed by almost two-thirds of respondents (63%). This approach was considered "very effective" by 24% of respondents. BI software has progressed from best-in-class niche products to BI platforms that provide integrated reporting, analysis, visualization, and dashboarding capabilities within a single, integrated architecture. In addition, many BI vendors now offer planning, budgeting, and consolidation applications to supplement their BI offerings.

WHAT STRATEGIES HAVE YOU EMPLOYED TO REMEDY THE PROBLEMS CAUSED BY SPREADMARTS, AND HOW EFFECTIVE WERE THEY?

	Employed?	VE?	NVE?
We leave them alone	71%	9%	58%
Provide a more robust BI/DW solution	63%	24%	21%
Create a set of standard reports for decision making	58%	18%	22%
Provide BI tools that tightly integrate with Excel/Office	53%	29%	21%
Stop providing IT support for spreadmarts	41%	6%	55%
Show executives how spreadmarts undermine compliance	34%	16%	40%
Create policies for the proper use of spreadsheets	18%	13%	50%
Let IT manage the spreadmart centrally	13%	13%	63%
Have executives issue a mandate against spreadmarts	13%	6%	64%

VE = very effective
NVE = not very effective

Table 1. Respondents could select more than one response.

We recommend caution with these BI replacement approaches. First, don't assume that business users will find the BI tools easy to use. Second, don't assume that business users will see the benefit of these systems if their spreadmarts are answering their business questions today. Get business users (not just power users) involved in the selection and implementation of BI tools, provide ongoing training, and market the benefits. "If it ain't broke, don't fix it"—if the business users are not committed to using the BI tools, walk away from the project and look for other spreadmarts the business perceives as a problem.

Create a Standard Set of Reports. Almost as many companies (58%) assumed that creating a standard set of reports using their standard BI tools would eliminate the need for spreadmarts as those that implemented new BI tools (63%). Organizations assumed that these reports would become their systems of record for decision making. Only 18% found this approach very effective. The most likely reasons for the shortcoming were, first, that no set of reports will effectively cover every management decision, so there was a gap in what was provided. Second, since this approach burdened IT with a queue of reports to develop, the business faced two of the primary reasons spreadmarts were created initially: the IT group did not understand what the business needed, and the IT group was not responsive to business needs.

Excel Integration. The only approach respondents rated more effective than adopting BI tools was "providing BI tools that integrate with Excel/Office" (29%). For a spreadmart user, the next best thing to Excel is Excel that integrates with the corporate BI standard. This approach was used by slightly more than half of the respondents (53%). However, Office integration technology can also provide users more fuel to proliferate spreadmarts if it enables users to save data locally and disseminate the results to users. Some BI vendors—and ironically, Microsoft is one of them—now provide a thin-client Excel solution where administrators can deny users the ability to download or manipulate data.

Some experts claim that power users use BI tools mainly as a personalized extract tool to dump data into Excel, where they perform their real work. According to our survey, that's not the case. Only a small percentage (7%) of spreadmarts obtain data this way. More than half of spreadmarts (51%) use manual data entry or manual data import. It follows that a major way to drain the life out of spreadmarts is to begin collecting the data they use in a data warehouse and create standard reports that run against that data. Of course, if there are no operational systems capturing this data, then a spreadmart is the only alternative.

Sometimes strong-arm tactics are effective in addressing spreadmarts. Reassigning the creators of spreadmarts to other activities is certainly effective, if an executive has the clout to carry this out and offers a suitable BI/DW replacement system. For example, the director of operations at a major national bank reassigned 58 people who were creating ad hoc performance reports with a set of standard reports created using a standard BI platform, saving $300 million a year and dramatically improving the bank's quality and efficiency in industry benchmarks. This may be the dream of those who are hostile to spreadmarts, but the survey illustrates that this is a rare occurrence.

Gentler approaches are seldom very effective. New policies for the proper use of spreadsheets generally fall on deaf ears; they are very effective only 12% of the time. The problem isn't that business people do not know how to use the spreadsheets, but that they think they have no alternative.

Multiple Solutions. Given the low percentage of respondents who can vouch for the effectiveness of any of the approaches listed in Table 1, it's not surprising that managing the proliferation of spreadmarts is such a difficult task. It is more of a change management issue than a technological one. While it's important to bring new technologies to bear, such as BI tools that integrate with Excel, it's critical to figure out which levers to push and pull to change people's habits and perceptions. No single approach is effective on its own; therefore, organizations must apply multiple approaches.

Conclusion

Spreadsheets are here to stay. Business users have them, are familiar with them, and will use them to do their jobs for years to come. Memo to IT: Deal with it! Our recommendation is to choose a solution that balances business and IT priorities and yields the greatest business value. ●

Wayne W. Eckerson is the director of TDWI Research at The Data Warehousing Institute. Eckerson is an industry analyst and the author of *Performance Dashboards: Measuring, Monitoring, and Managing Your Business* (John Wiley & Sons, 2005). He can be reached at weckerson@tdwi.org.

Richard P. Sherman is the founder of Athena IT Solutions, a Boston-area firm offering data warehousing and business intelligence consulting and training. He is an expert instructor and speaker at industry conferences and seminars and teaches at Northeastern University's graduate school of engineering. He can be reached at rsherman@athena-solutions.com.

This article was excerpted from the full, 28-page report by the same name. You can download this and other TDWI Research free of charge at www.tdwi.org/research/reportseries.

The report was sponsored by Actuate, Cognos, Microsoft, MicroStrategy, Pentaho, SAP, Unisys, and XLCubed.

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, an SAP company

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

As an independent business unit within SAP, Business Objects transforms the way the world works by connecting people, information, and businesses. Together with one of the industry's strongest and most diverse partner networks, the company delivers business performance optimization to customers worldwide across all major industries, including financial services, retail, consumer-packaged goods, healthcare and public sector. With open, heterogeneous applications in the areas of governance, risk and compliance; enterprise performance management; and business intelligence; and through global consulting and education services, Business Objects enables organizations of all sizes around the globe to close the loop between business strategy and execution.

Collaborative Consulting

70 Blanchard Road, Ste. 500
Burlington, MA 01803

781.565.2600
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 can 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 Corporation, a SAS Company

940 NW Cary Parkway, Ste. 201
Cary, NC 27513

877.846.3589
Fax: 919.447.3100

sales@dataflux.com
www.dataflux.com

DataFlux enables organizations to analyze, improve, and control their data through an integrated technology platform. With DataFlux enterprise data quality and data integration products, organizations can more effectively and efficiently build a unified view of customers, products, suppliers, or any other corporate data asset. A wholly owned subsidiary of SAS (www.sas.com), DataFlux helps customers rapidly assess and improve problematic data and build the foundation for enterprise data governance. Effective data governance delivers high-quality information that can fuel successful enterprise efforts such as risk management, operational efficiency, and master data management (MDM). To learn more, visit www.DataFlux.com.

DATAllegro, Inc.

85 Enterprise, 2nd Floor
Aliso Viejo, CA 92656

949.330.7690
Fax: 949.330.7691

info@datallegro.com
www.datallegro.com

DATAllegro v3™ is the industry's most advanced data warehouse appliance utilizing an all-commodity platform. By combining DATAllegro's patent-pending software with the industry's leading hardware, storage, and database technologies, DATAllegro has taken data warehouse performance, reliability, and innovation to the next level. DATAllegro 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, DATAllegro'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.

Dataupia Corporation

One Alewife Center
Cambridge, MA 02140

866.748.DATA
617.301.8500
Fax: 617.441.7776

info@dataupia.com
www.dataupia.com

Dataupia, founded in 2005, brings a strong record of industry leadership to addressing the growing gap between the massive volumes of stored data and the portion that a business can use to its benefit.

By architecting specialized software and industry-standard hardware into a highly cost-effective and intelligent data warehouse appliance, Dataupia's solution will amplify an organization's existing information systems to provide deeper access into their data universe and more comprehensive business insight. In recognition of its omniversal transparency™, ease of integration, scalability and performance, the Dataupia™ Satori Server was selected as one of SearchDataManagement's "2007 Products of the Year."

Learn more at www.dataupia.com.

IBM Corporation

1 New Orchard Road
Armonk, NY 10504-1722

1.800.IBM.4YOU

www.ibm.com/contact/us/
www.ibm/software/data/integration

As the world's largest information technology company, IBM has 80+ years of leadership in helping business innovate. IBM delivers market-leading solutions to critical information-intensive business problems, allowing customers to achieve new levels of innovation through best-of-breed information management capabilities. IBM has provided customers with hardware, software, and services solutions over the years, helping customers leverage their information. IBM Information Server helps organizations derive more value from the complex, heterogeneous information spread across their systems.

Identity Systems

1445 East Putnam Avenue
Old Greenwich, CT 06870

203.698.2399
Fax: 203.698.2409

USASales@identitysystems.com
www.identitysystems.com

Identity Systems is a global leader in enterprise software used for data quality, master data management, business intelligence, and identity resolution. Identity Systems develops fast, highly accurate, and scalable solutions to profile, cleanse, group, match, and merge data within computer systems and across network databases. Founded in 1986, Identity Systems has over 600 clients worldwide in commercial and governmental organizations, including the U.S. Department of Homeland Security, IRS, Florida Department of Law Enforcement, State of Texas, GE Money, Citigroup, VISA, American Express, Equifax, Experian, British Telecom, Zurich Financial, The Hartford, Kaiser Permanente, HP, AT&T, Sprint, and Federal Express.

Information Builders

Two Penn Plaza
New York, NY 10121-2898

212.736.4433
Fax: 212.967.6406

askinfo@ibi.com
www.informationbuilders.com

Information Builders' award-winning combination of business intelligence and enterprise integration software has been providing innovative solutions to more than 12,000 customers for the past 30 years. WebFOCUS is the world's most widely utilized business intelligence platform. It provides the security, scalability, and flexibility needed at every level of global extended enterprises.

iWay Software suite provides state-of-the-art, multipurpose, prebuilt integration components that address all SOA, application, data, and information management requirements. Its integration adapters have been adopted by the leading software platform providers. Together, these products give Information Builders' customers the ability to live up to the company motto: *Your business. No barriers.*

Headquartered in New York City with 90 offices worldwide, the company employs 1,450 people and has more than 350 business partners.

Initiate Systems, Inc.

200 West Madison, Ste. 2300
Chicago, IL 60606

312.759.5030
Fax: 312.759.5026

info@InitiateSystems.com
www.InitiateSystems.com

Initiate Systems, Inc. enables organizations to strategically leverage and share critical data assets. Its master data management (MDM) software and its experience as an information exchange leader provide organizations with complete, accurate, and real-time views of data spread across multiple systems or databases, even outside the firewall. This allows companies to unlock the value of their data assets for competitive advantages or operational improvements. Initiate Systems operates globally through its subsidiaries, with corporate headquarters in Chicago and offices across the U.S. and in Toronto, London, and Sydney. For more information, visit www.InitiateSystems.com.

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 thousands of customer successes and a reputation for innovation and leadership, MicroStrategy is the clear choice for your business intelligence requirements. More information about MicroStrategy is available at www.microstrategy.com.

SenSage, Inc.

55 Hawthorne Street, Ste. 700
San Francisco, CA 94105

415.808.5900
Fax: 415.371.1385

info@sensage.com
www.sensage.com

SenSage, Inc. offers the only patented event data warehousing solution for compliance regulations, fraud detection, and analytical uses of event data. More than 300 customers have deployed SenSage solutions, including many of the leading telecommunications providers for compliance with the EU Data Retention Directive known as Directive 2006/24/EC. Based in San Francisco, the company markets its solutions directly and through partners, including Cerner, EMC, HP, Hitachi Data Systems, IBM, McAfee, Tokyo Electron Devices, and many others. Visit www.SenSage.com for more information.

Siperian, Inc.

1820 Gateway Drive, Ste. 109
San Mateo, CA 94404

650.571.2200
Fax: 650.350.2206

sales@siperian.com
www.siperian.com

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

Syncsort Incorporated

50 Tice Boulevard
Woodcliff Lake, NJ 07677

201.930.8200

marcom@syncsort.com
www.syncsort.com

Syncsort is a leading developer of high-performance data management and data warehousing software. For nearly 40 years, Syncsort has built a reputation for superior product performance and reliable technical support. Most of the *Fortune* 500 companies are Syncsort customers, and Syncsort's products are used in more than 50 countries to speed data warehouse processing and improve performance of data-intensive applications and processes.

DMExpress is Syncsort's high-speed data integration tool. It speeds large-volume applications, saving customers hours—even days—of processing time.

Talend

105 Fremont Avenue, Ste. F
Los Altos, CA 94022

714.786.8140
Fax: 714.786.8139

info@talend.com
www.talend.com

Talend is the first provider of open source data integration software. After three years of intense research and development investment, and with solid financial backing from leading investment firms, Talend revolutionized the world of data integration when it released the first version of Talend Open Studio in 2006.

Talend's solutions are used primarily for integration between operational systems, as well as for ETL (extract, transform, load) for business intelligence and data warehousing, and for migration.

Unlike proprietary, closed solutions, which can only be afforded by the largest and wealthiest organizations, Talend makes data integration solutions available to organizations of all sizes and for all integration needs.

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).

tdwi
THE DATA WAREHOUSING INSTITUTE

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 7,000 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.

tdwi PARTNER

baseline CONSULTING

Business Objects an SAP company

COGNOS AN IBM COMPANY

CONNECT: The Knowledge Network

DataFlux A SAS COMPANY

DATAllegro DATA AT THE SPEED OF BUSINESS

DATAupia™ Free your data

DecisionPath CONSULTING

hp invent

IBM

identitysystems

INFORMATICA The Data Integration Company™

kognitio Competitive advantage from data

Microsoft

MicroStrategy Best In Business Intelligence

NETEZZA The Power to Question Everything™

ORACLE

Pitney Bowes GROUP 1 SOFTWARE

sas

syncsort

TERADATA Raising Intelligence