POWERFUL CASE STUDIES AND
LESSONS LEARNED FOCUSING ON:
Data Integration
Data Quality
Master Data Management
Open Source Data Integration
Data Warehousing

FEATURE
Collaborative Data Integration
Philip Russom, TDWI Research
Find out why collaboration requirements for data integration projects have intensified this decade.
PAGE 2

SPECIAL SECTION: DATA WAREHOUSING
Redefining the Data Warehouse Appliance
Philip Russom, TDWI Research
Learn about some of the trends and developments that are driving a redefinition of the data warehouse appliance.
PAGE 30

TDWI RESEARCH EXCERPTS
Performance Management Strategies: Understanding KPIs
PAGE 36

Introduction to Operational Data Integration
PAGE 41
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 and data warehousing today. We’ve arranged these case studies and lessons from the experts into specific categories to guide you through the articles: data integration, data quality, master data management, and open source data integration, as well as a special section on data warehousing.

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 data integration technologies and solutions for industries ranging from credit unions to schools of medicine to global benchmarking tracking services.

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 data integration best practices and trends.

FEATURE ARTICLE
In this issue, the feature article comes from Philip Russom, senior manager of TDWI Research. In “Collaborative Data Integration,” he discusses how collaboration requirements for data integration projects have intensified this decade, mainly due to the increasing number of data integration specialists within organizations, the geographic dispersion of data integration teams, and the need for business people to perform stewardship for data integration.

SPECIAL SECTION: DATA WAREHOUSING
This issue of What Works also features a special section devoted to data warehousing. In this section, you will find more case studies and lessons from the experts, as well as “Redefining the Data Warehouse Appliance” by Philip Russom. In this piece, he examines some of the trends and developments that are driving a redefinition in the data warehouse appliance.

TDWI RESEARCH
There’s more from TDWI Research. What Works includes excerpts from TDWI’s recent best practices reports: Performance Management Strategies: How to Create and Deploy Effective Metrics, TDWI’s latest report from Wayne Eckerson, and Operational Data Integration: A New Frontier for Data Management from Philip Russom.

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

FEATURE
2 Collaborative Data Integration
Philip Russom, Senior Manager, TDWI Research
Find out why collaboration requirements for data integration projects have intensified this decade.

CASE STUDIES AND LESSONS FROM THE EXPERTS
6 Data Integration and Data Warehousing Defined

DATA INTEGRATION
8 Key Considerations When Implementing Real-Time Data Integration Solutions
9 Empowering Nationwide Agents with Dashboards and Reports to Reach Sales Targets
10 Evolving from Disparate, Departmental BI to a Fully Integrated Enterprise BI Environment
11 UCHC Leverages Enterprise Job Scheduling to Automate Data Integration Processes
12 Why End-to-End ETL Automation Is a Must for BI Success
13 Visionary School District Finds New Ways to Eliminate IT Interoperability Issues
14 Advantages of a Software-Based Data Sharing Solution

DATA QUALITY
15 Nectar Card: The Data Behind the Brand
16 Demanding Enterprisewide Data Quality
17 PITT OHIO EXPRESS Delivers Results with Improved Customer Data
18 With Data, the Whole Is More than the Sum of Its Parts
19 Total Data Quality—Just What the Doctor Ordered
20 Finally the Right Tools for the Job: Fix the Flat Tire, Replace the Plugs, or Buy a New Car
21 New EIM Requirements Demand Better Solutions

MASTER DATA MANAGEMENT
23 Selling MDM to the C-Suite
24 Multidomain Master Data Management
25 MDM Brings Financial Services Closer to Customers
26 Righteous Data: The Business Value of Data Integration
27 Compliance and Risk Management: Ten Critical MDM Capabilities to Include in Your Evaluation

OPEN SOURCE DATA INTEGRATION
29 Prime Numbers Technology Revolutionizes the Analysis of Corporate Travel Spending with Open Source Data Integration

SPECIAL SECTION: DATA WAREHOUSING
30 Redefining the Data Warehouse Appliance
Philip Russom, Senior Manager, TDWI Research
Learn about some of the trends and developments that are driving a redefinition of the data warehouse appliance.
34 Revol Wireless Dials in Key Company Strategies, Competitive Advantages with Business Intelligence Solutions
35 Best Practices for Architecting Your Data Warehouse

TDWI RESEARCH: BEST PRACTICES REPORTS EXCERPTS
36 Performance Management Strategies: Understanding KPIs
41 Introduction to Operational Data Integration

MORE INFORMATION
45 Solution Providers
51 About TDWI
52 TDWI Partner Members
Collaboration requirements for data integration projects have intensified greatly this decade, largely due to the increasing number of data integration specialists within organizations, the geographic dispersion of data integration teams, and the need for business people to perform stewardship for data integration. Organizations experiencing these trends need to build teams, best practices, and infrastructure for the emerging practice known as **collaborative data integration**.

TDWI Research defines collaborative data integration as:

A collection of user best practices, software tool functions, and cross-functional project workflows that foster collaboration among the growing number of technical and business people involved in data integration projects and initiatives.

In May 2007, TDWI asked conference attendees a few questions about collaborative data integration to find out if they were aware of it and to quantify a few aspects of its practice. Judging by survey responses, data management professionals of different types are clearly aware of the practice. In fact, almost two-thirds of survey respondents reported that collaboration is an issue for data integration in their organizations. (See Figure 1, next page.)

**Why You Should Care about Collaborative Data Integration**

Several trends are driving up the requirements for collaboration in data integration projects:

- **Data integration specialists are growing in number.** Collaboration requirements intensify as the number of data integration specialists increases. Many organizations have moved from one or
Data governance and other forms of oversight touch data integration.

• **Data integration specialists are expanding their work beyond data warehousing.** Analytic data integration focuses mainly on data warehousing and similar practices like customer data integration (CDI). This established practice is now joined by operational data integration, which focuses on the migration, consolidation, and upgrade of operational databases. Both practices are growing and thereby increasing personnel.

• **Data integration work is increasingly dispersed geographically.** Projects that involve data integration are progressively outsourced, which demands procedures and infrastructure for communication among internal resources and external consultants. Even when the entire project team works for the same organization, employees may work from home, from various offices, or while traveling.

• **Data integration is now better coordinated with other data management disciplines.** Data integration specialists must coordinate efforts with specialists for data quality, metadata management, data warehousing, master data management, operational applications, database administration, and so on. These specialists all experience moments where they must work together or simply have a read-only view of data integration project artifacts. Related to this, multidisciplinary data management is coordinated more and more by a central enterprise data architect who authorizes the design of, and enforces standards for, data integration work.

• **More business people are getting their hands on data integration.** Stewardship for data quality has set a successful precedent. Inspired by that model, a few bold business folks are browsing the repositories of data integration tools to identify data that needs integration and track the progress of integration work that they’ve commissioned or sponsored. This form of collaboration ensures that data integration truly supports the needs of a business.

• **Data governance and other forms of oversight touch data integration.** Data is so central to several compliance requirements that governing bodies need to look into data integration projects to understand whether they are compliant. For example, one of the typical responsibilities of a data governance committee is to ensure that data for regulatory reports is drawn from the best sources and documented with an auditable paper trail. Goals like this are achieved faster and more accurately when supported by the “big picture” that collaborative data integration provides.

In summary, the number and diversity of people involved in data integration planning and execution are increasing, thus demanding better practices and software tools for collaboration.

### Organizational Issues with Collaborative Data Integration

**The Scope of Collaboration for Data Integration**

With data integration initiatives, the scope of collaboration varies, depending on the variety of organizational units involved.

Sometimes collaboration focuses narrowly on technical people who develop data integration implementations, but it may also include people from IT management who oversee data integration work, like a BI director or enterprise data architect. When data integration work is outsourced, communication between the client company and consultants is another form of collaborative data integration. With analytic data integration, the data integration team must collaborate with business analysts and report producers to ensure they have the data they need. Of special note, the scope of collaboration is progressively extending to business people whose success relies on integrated data, ranging from the chief financial officer to line-of-business managers.

All these folks together constitute an extended, cross-functional team of broad scope, held together by common goals like compliance, quality decision making, customer service, information improvement, or leveraging data as an enterprise asset. This diverse team needs a centralized organizational structure and technology infrastructure through which every team member can contribute to the business alignment, initial planning, tactical management, and project implementation of initiatives for data integration and related data management techniques.

### Organizational Structures that Make Data Integration Collaborative

Different organizational units provide a structure in which data integration can be collaborative:

- **Technology-focused organizational structures.** Data integration—especially when it’s for operational purposes, not analytic ones—is sometimes executed by a data management group. The focus is on technology implementations and administration, though with guidance from business sponsors. More and more, data integration is being commissioned by an enterprise...
data architecture group, which TDWI sees as a new evolution beyond data management groups. In these cases, the scope of collaboration is mostly among technical workers.

- **Business-driven organizational structures.** At the other end of the spectrum, people focused on business opportunities determine where data integration and related techniques (such as data quality, master data management, metadata management, and so on) can serve grander business goals. Examples include data stewardship programs (especially when they expand beyond data quality to encompass data integration), data governance committees (which govern many data management techniques, not just data integration), and steering committees (although these tend to be temporary). The scope of collaboration is very broad, covering business people (who initiate data integration projects, based on enterprise needs), technical people (who design and execute implementations), and stewards and project managers (who are hybrid liaisons).

- **Hybrid structures.** Some organizations align business and technology by nature, and hence are hybrids. For example, the average BI or data warehousing team seems focused on technology, due to the technical rigor required for data integration, warehouse modeling, report design, and so on. Yet most team members communicate regularly with sponsors, report consumers, and other business people to ensure that information delivery and business analysis requirements are met. Because most data integration competency centers are spawned from data warehousing teams, they too are usually hybrids. In these cases, the scope of collaboration reaches across business and IT people, who work on fairly equal ground.

Corporations and other user organizations have hired more in-house data integration specialists in response to an increase in the amount of data warehousing work and operational data integration work outside of warehousing. In the “old days,” an organization had one or maybe two data integration specialists in-house, whereas today’s average is closer to three or four. In fact, roughly a quarter of respondents to a TDWI survey reported five or more, while another quarter reported three or four. (See Figure 2.)

**Software Tools for Collaborative Data Integration**

Although much of the collaboration around data integration consists of verbal communication, software tools for data integration include functions that automate some aspects of collaboration. Tools are increasingly providing functions through which nontechnical and mildly technical people can collaborate. However, software tool automation today mostly enables collaboration among data integration specialists who design, develop, deploy, and administer data integration implementations. The focus on collaborative implementation is natural, since it’s driven by the need to support the large data integration teams that evolved early this decade.

Data Integration Tool Requirements for Technical Collaboration

Let’s recall that the mechanics of collaboration involve a lot of change management. It trickles down all the way to data integration development artifacts, such as projects, objects, data flows, routines, jobs, and so on. Managing change at this detailed level—especially when multiple developers (whether in-house or outsourced) handle the same development artifacts—demands source code management features in the development environment. Again, these features have existed in other application development tools, but were only recently added to data integration tools:

- **Check out and check in.** The tool should support locking for checked-out artifacts and automated versioning for checked-in ones. Role-based security should control both read and write access to artifacts.

- **Versioning.** This should apply to both individual objects and collections of them (i.e., projects). The tool should keep a history of versions with the ability to roll back to a prior version and to compare versions of the same object. This functionality may be included in the data integration tool or provided by a more advanced third-party version-control system.

- **Nice-to-have source code management features.** Check in/out and versioning are absolute requirements. Optional features include project management, project progress reports, object annotation, and discussion threads.

Data Integration Tool Requirements for Business Collaboration

Development aside, a few data integration and data quality tools today support areas within the tools for data stewards or business personnel to use. In such an area, the user may actively do some hands-on work, like select data structures that need quality or...
integration attention, design a rudimentary data flow (which a technical worker will flesh out later), or annotate development artifacts (e.g., with descriptions of what the data represents to the business).

However, most collaboration coming from stewards, business analysts, and other business people is more passive. Therefore, some data integration tools provide views of data integration artifacts that are meaningful to these people, like business-friendly descriptions of semantic layers and data flows. Other views may focus on project documents that describe requirements, staffing, or change management proposals. Depending on what’s being viewed, the user may be able to annotate objects with questions or comments. Some tools support discussion threads to which any user can contribute.

All of these functions are key to extending collaboration beyond the implementation team to other, less technical parties. Equally important, however, is that different tools (or areas within tools) bridge the gaps among diverse business and technical user constituencies. Such bridges are best built via common data integration infrastructure with a shared repository at its heart.

Collaboration via a Tool Depends on a Central Repository

The views just described are enabled by a repository that accompanies the data integration tool. Depending on the tool brand, the repository may be a dedicated metadata or source code repository that has been extended to manage much more than metadata and development artifacts, or it may be a general database management system. Either way, this kind of repository manages a wide variety of development artifacts, semantic data, project documents, and collaborative views. When analytic data integration is applied to business intelligence and data warehousing, the repository may also manage objects for reports, data analyses, and data models.

So that all collaborators can reach it, the repository should be server based and easily accessed over LAN, WAN, and Internet. Depending upon the user and what he/she needs to do, access may be available via the data integration tool or a Web browser, but with security. A well-developed repository will support views (some read only, others with write access) that are appropriate to various collaborators, including technical implementers, IT management, business people, business analysts, data stewards, BI professionals, and data governance folks. Because the repository manages development artifacts and semantic data, it may also be accessed by the data integration tool as it runs regularly scheduled jobs. Hence, the repository is a kind of hub that enables both the daily operation of the data integration tool and broad collaboration among all members of the extended data integration team.

Recommendations

- **Recognize that data integration has collaborative requirements.** The greater the number of data integration specialists and people who work closely with them, the greater the need is for collaboration around data integration. Head count aside, the need is also driven up by the geographic dispersion of team members, as well as new requirements for regulatory compliance and data governance.

- **Determine an appropriate scope for collaboration.** At the low end, bug fixes don’t merit much collaboration; at the top end, business transformation events require the most.

- **Support collaboration with organizational structures.** These can be technology focused (like data management groups), business driven (data stewardship and governance), or a hybrid of the two (BI teams and competency centers). Organizational units such as these are best led by dual chairs representing IT and the business.

- **Select data integration tools that support broad collaboration.** For technical implementers, this means data integration tools with source code management features (especially for versioning). For business collaboration, it means an area within a data integration tool where the user can select data structures and design rudimentary process flows for data integration. Equally important, the tool must bridge the gap so that planning, documentation, and implementations pass seamlessly between the two constituencies.

- **Demand a central repository.** Both technical and business team members—and their management—benefit from an easily accessed, server-based repository through which everyone can share their thoughts and documents, as well as view project information and semantic data relevant to data integration.

**Philip Russon** is the senior manager of TDWI Research at The Data Warehousing Institute (TDWI), where he oversees many of TDWI’s research-oriented publications, services, and events. He’s been an industry analyst researching BI issues at Forrester Research, Giga Information Group, and Hurwitz Group. You can reach him at prussom@tdwi.org.

This article was excerpted from the TDWI Monographs *Collaborative Data Integration: Coordinating Efforts within Teams and Beyond* and Second-Generation Collaborative Data Integration: Sustainable, Operational, and Governable. Both are available online at www.tdwi.org/research/monographs.
Fundamentally, data warehousing is an exercise in data integration. A data warehouse attempts to reintegrate 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 cleansed, 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 quality is a complex concept that encompasses many data management techniques and business quality practices, applied repeatedly over time as the state of quality evolves, to achieve levels of quality that vary per data type and seldom aspire to perfection. The most common technique is name-and-address cleansing, whereas the least common is the internationalization of data for quality purposes. Between these two extremes are numerous data quality techniques, including data standardization, verification, profiling, monitoring, matching, merging, householding, geocoding, postal standards, enrichment, and so on.

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.

Open source data integration tools are, for the most part, like other data integration tools in terms of functionality. One difference is that, being based on open source code, the functionality is developed primarily by a software vendor, but augmented by the efforts of a developer community. Another difference is that most are not licensed to users per se; instead, most users download and use the open source tool at no charge, then pay a minimal charge for support and maintenance later, if they choose to deploy a solution. Open source data integration tools are known for their low cost, ease of customization, quick procurement via download, and ease of embedding within various applications.

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 metadata repository. Tools that support these various modeling, mapping, and documentation activities are known as data warehouse design tools.
The New York Independent System Operator (NYISO) is the not-for-profit corporation responsible for operating New York’s bulk electricity grid, administering the state’s competitive wholesale electricity markets, and conducting comprehensive long-term planning for its electric power system.

The NYISO manages the efficient flow of power on more than 10,800 miles of high-voltage transmission lines on a minute-to-minute basis, 24 hours a day, every day of the year. As the administrator of competitive wholesale electricity markets, it also conducts auctions that match the buyers and sellers of power.

Challenge

The NYISO is dedicated to building and maintaining trust in its operation of the bulk electricity grid, in the information it provides, and in its role as the independent, impartial administrator of the state’s electricity markets. The NYISO offers a variety of marketplace information, such as pricing data, to help those who buy and sell electricity in New York’s open, fair, and competitive energy markets.

At the forefront is the Comprehensive Bid Management System—a cutting-edge, Web-based platform where more than $70 billion in transactions have been made since 1999. The NYISO’s Price Validation Group performs a high rate of analysis, using the system to track the pricing, bidding, and purchasing of electricity that takes place every day in New York State, and also to ensure that the market is functioning properly.

Previously, analysts had to run reports against the production system, which could negatively impact system performance. Thus, the team was restricted to running small queries. Alternatively, the group would have to wait for the nightly batch loads to reach the reporting server before it could access information for analysis. This could cause delays in uncovering potential issues.

The IT team at the NYISO began its search for a technology solution that would move its mission-critical data in real time, from the organization’s production system running on Oracle 9i to an operational data store (ODS) running on Oracle 10g, for reporting, analysis, and quicker resolution of any market discrepancies. The NYISO IT team set out to find a partner that could meet all of its technical requirements, including the ability to handle high volumes and scale over time, low-latency data movement, minimal intrusiveness on the production system, and the ability to support its various hardware environments.

Solution

The NYISO performed a proof of concept (POC) with GoldenGate Software. “The team was very seasoned and was able to immediately troubleshoot and remedy any issues we encountered, while streamlining the replication process and making the POC a complete success—ultimately sealing the deal to move forward with GoldenGate,” said Michele Armstrong, senior database administrator for the NYISO.

Access to the real-time data allowed the NYISO’s Price Validation Group to review transactions quickly and take action accordingly.

Shortly thereafter, the NYISO was live with the GoldenGate real-time data integration solution for live reporting to capture transactional data from the source application, the Comprehensive Bid Management System, and streaming it in real time to a secondary Oracle operational data store for look-ups, reporting, and analytics. Access to the real-time data not only allowed the NYISO’s price validation group to review transactions quickly and take action accordingly, but it also allowed the IT team at the NYISO to automate certain alerts and processes to improve overall efficiency and make life even easier for the analysts.

John Bub, product manager for the NYISO’s business intelligence platform, concluded: “The software solution that GoldenGate offers significantly impacts our business. It has delivered value quickly and we look forward to seeing additional value as we roll it out for other projects across the organization.”

For a free white paper on this topic, click here and choose the title “Real-Time Data Integration for Data Warehousing and Operational BI.” For more information about GoldenGate Software, click here.
Key Considerations When Implementing Real-Time Data Integration Solutions

By Chris McAllister
Senior Director of Product Management, GoldenGate Software

For years, companies have relied on the integration of data to measure results and improve strategic planning. Forward-thinking companies, however, are not only looking at the past to plan the future, but are also looking at what is happening right now to influence current, dynamic business activity. Data can now be leveraged to empower operational decision making, to answer questions such as “Which products will need to be restocked in the next two hours?” or “Which customers making purchases right now would be likely to accept a cross-sell offer?” That means putting data into the hands of front-line employees across the enterprise, such as call centers, the manufacturing floor, and store operations.

A crucial requirement is the ability to integrate—and provide access to—real-time or near-real-time operational data. There are several approaches and technologies that may be evaluated to achieve real-time data integration, but they differ in important ways. Companies should make the following considerations part of their evaluation checklist.

Data Timeliness and IT Impact
Understanding the data latency requirements of the business is an important factor. Some may need only hourly data, while others may need data to be delivered in minutes or seconds. The IT team should evaluate the level of complexity in implementing mini batches or intraday batches versus designing for continuous real-time feeds. They may discover that four-times-a-day or hourly batch architectures are far more complicated to implement and also impose too much impact throughout the business day. Further, a longer-term view should be taken: Hourly updates today may mean real-time updates will be requested next year. Continuous data feeds can be implemented with negligible system impact; architecting a solution where technology isn’t the constraint may save time and money in the long run.

Data Volumes and Performance
The optimal data integration solution should be capable of keeping up with the volume of new and changed data at the determined latency. In many cases, this could mean moving thousands of transaction operations per second. Spikes during peak times should also be handled while adding little or no additional latency.

Batch-Independent Processing
In a 24/7 world, batch windows are shrinking rapidly. Higher data volumes combined with more frequent integration make batch-oriented technologies a limiting choice when moving toward a lower-latency architecture. The optimal data integration solution should not rely on a batch window.

ETL Integration
The most common method for acquiring data from source systems is to use extract, transform, and load (ETL) tools, but ETL tools process data in batches and OLTP activity must cease when they do so. However, the transformation and data cleansing capabilities may be necessary for some subsets of the data. A real-time data capture solution should be able to integrate easily with ETL tools when more extensive transformation and cleansing are required.

Flexibility
A data integration solution by definition should support a variety of heterogeneous environments, but beyond that it should support myriad topologies and be easy to implement, manage, and scale as needs change. For example, adding new data sources and/or targets should be straightforward and not require major overhauls.

Data Integrity and Recoverability
When moving and integrating data at the transaction level, the solution should maintain the referential integrity of each transaction between source and target systems. The architecture should also be built to easily recover from unexpected interruptions, such as hardware failures, network issues, and human errors, without losing or corrupting the transactional data.

Conclusion
Real-time data integration solutions are getting a lot of attention, but they are not all alike. Unless the solution delivers real-time data without compromising the performance of OLTP systems, and unless this information is continuously available, organizations will not be able to realize the full benefits of leveraging operational data across the enterprise.

Continuous feeds can be designed for negligible impact—and may save time and money in the long run.
Empowering Nationwide Agents with Dashboards and Reports to Reach Sales Targets

Commentary by Lori Grose
Project Lead, Nationwide Insurance
and Ravi Ranjan
Consulting IT Architect, Nationwide Insurance

Nationwide Insurance has successfully operated for more than 80 years, with more than 3,500 exclusive agencies across the country. In recent years, the property and casualty side of Nationwide, which brings in close to 80 percent of its overall business, needed better access to information to reach its sales targets.

In 2006, Nationwide chose the Teradata Data Warehouse Platform to create a single data source that would consolidate all sales and policy information. In addition, Nationwide selected the MicroStrategy Business Intelligence Platform to integrate dynamic enterprise dashboards into its reporting system. This made it easy for agents to view sales activity at a glance.

In 2008, Nationwide launched Revenue Connection, one of the organization’s most mission-critical implementation projects of the year. This BI application monitors the degree to which Nationwide’s property and casualty team is achieving its key strategic objectives and goals. Specifically, end users now have access to the policy information they need to achieve their sales goals and run their business.

Dynamic Enterprise Dashboards Support Informed Decisions

Revenue Connection delivers easy-to-use dashboards and reports to Nationwide’s agents and field management. End users can access visually engaging dashboards that incorporate Adobe Flash and advanced visualization widgets using Adobe Flex Builder.

End users can see graphical displays of data with prior year comparisons, data in year-to-date format, or 12-month moving formats to determine trends. In addition, they can access product or policy details previously not available.

Currently there are 100 reports and more than 30 dashboards in Revenue Connection. This level of innovation, using state-of-the-art dashboard technology, demonstrates Nationwide’s desire to stay ahead of its competitors.

The ETL layer of Revenue Connection, also from Teradata, sources data from approximately 30 interfaces and then standardizes, integrates, and loads the data into approximately 60 tables. The parallel architecture of Teradata enables the ETL process to load millions of transactions in a couple of hours. Currently the database size is 750 GB, and new data sources are in the process of being added to Revenue Connection.

The reporting layer, built using MicroStrategy, uses approximately 120 views to report the data to nearly 10,000 end users. These views utilize the unique capability of Aggregate Join Index to provide faster response times. The capability of MicroStrategy to run against large databases enabled the team to achieve aggressive performance results. Dashboard response time is 10–20 seconds and the reports run in 4 to 45 seconds, depending upon the amount of data. While many other business units have attempted to create dashboards and scorecards, no one within Nationwide achieved the performance results meeting these complex dashboard requirements.

For a free white paper on this topic, click here and choose the title “MicroStrategy 9: Breakthrough Technology that Supports the Full Range of Business Intelligence.” For more information about MicroStrategy, click here.
Evolving from Disparate, Departmental BI to a Fully Integrated Enterprise BI Environment

By Dan Paladino
Analyst, Industry and Solution Marketing, MicroStrategy

The ideal architecture for BI is an enterprise architecture in which there is a cohesive model of the business represented in a single metadata across all applications. Enterprise architecture helps ensure a single version of the truth by encompassing all data sources, reports, dashboards, and other BI analyses in an organization. It is also the most efficient architecture to own and operate, and incurs a low cost of ownership as a result.

Despite the clear advantages of enterprise BI, achieving it can be a challenge in terms of the political and IT effort it requires. In addition, businesses often cannot wait to achieve fully integrated enterprise BI; quick solutions are a must. Departmental BI islands provide them with fast and nimble BI. As a result, multiple departmental BI systems surface within an organization.

Despite the freedom and flexibility provided by islands of departmental BI, it is not an ideal or long-term solution for a company. As small, independently operated BI applications arise, it becomes inevitable that conflicting versions of the truth do so as well. Since each departmental system consists of its own reports and metadata accessing its own data sources, there is no data uniformity with other departmental or enterprise-grade applications.

The cost of maintaining these departmental systems poses another challenge to companies trying to save in a challenging economy. An organization’s resources are rarely shared efficiently, if at all, and the organization cannot benefit from the economies of scale.

To help ensure a single version of the truth and reduce costs, it is critically important to consolidate departmental islands into an enterprise BI environment. In order for an organization to evolve to an enterprise architecture, every departmental BI system in the IT landscape—including its metadata objects and data sources—will eventually need to be encompassed within a single, unified metadata.

Ideally, the first stage in the process of consolidating BI islands should not necessarily require the actual movement of data from departmental data sources to the enterprise data warehouse. Instead, the previously distinct metadata definitions should be merged gradually into a unified enterprise metadata that spans the vast collection of databases in the organization. This process must be incremental, with the metadata of each BI island consolidated into the enterprise one at a time. The next stage involves incrementally consolidating all of an organization’s data into an enterprise data warehouse.

However, it is crucial that these stages allow companies to gradually consolidate islands of BI without interrupting their business users. Departmental users should be able to continue to use their BI applications and reports even as they are consolidated to an enterprise architecture. In addition, an organization’s IT group should not be required to redesign any of those reports or other metadata objects because of the data consolidation. This transparency is an essential factor as departmental islands are consolidated.

Even within a larger, overarching enterprise framework, a BI technology must allow individual departments and workgroups to retain their departmental autonomy and nimbleness even as consolidation takes place. Users of departmental BI applications do not (and should not need to) sacrifice their convenient and flexible BI access as the organization works to achieve unified enterprise BI. The most sophisticated enterprise BI environments cater to the needs of both their enterprise and departmental business users.

Summary
Despite the nimbleness of departmental BI, organizations want to eventually achieve enterprise BI, which provides a single version of the truth, high efficiency and performance, and the lowest cost of ownership. This can only be achieved through a gradual, seamless consolidation of BI islands.

For a free white paper on this topic, click here and choose the title “MicroStrategy 9: Breakthrough Technology that Supports the Full Range of Business Intelligence.” For more information about MicroStrategy, click here.
UCHC Leverages Enterprise Job Scheduling to Automate Data Integration Processes

Commentary by Kevin McDonnell, Jay Coppola, and Bob Senack
Data Services Team, University of Connecticut Health Center

In 2006, the University of Connecticut Health Center (UCHC), a leading biomedical research center, healthcare facility, and medical school, launched a data warehousing initiative designed to leverage its massive pool of data.

Essentially, UCHC wanted and needed one source of truth. That requirement mandated creating a global data warehouse that would dissolve barriers between departments and produce uniform and accurate reporting across the organization.

The Data Integration Challenge

Prior to using enterprise scheduling technology, UCHC managed data extraction and loading manually through customized scripting, point solutions, and hands-on management. Though IT was usually able to deliver results within the time frames that users required, the effort was labor intensive and inefficient. In a typical scenario, the data warehouse team would run a data load that was dependent on certain files arriving via FTP. The team would have to check repeatedly to determine if the files had arrived and then run the load when they were all present. If any of the files were corrupted, it could take hours to recover from the resulting errors.

The Solution: Robust Job Management through Enterprise Scheduling

The IT team chose Tidal Enterprise Scheduler for its ability to automate the complex data integration process, manage dependencies, and ultimately ensure the accuracy of information and reporting.

The data warehouse team found that Tidal Enterprise Scheduler met its needs for a secure, easy-to-use interface and central location for monitoring all job processes. Tracking the movement of files from one group of users to another became faster and simpler. In addition, the group was able to assume responsibility for FTP jobs and data duplication tasks that had been scattered across other UCHC departments and manage them more efficiently. Tidal’s method of handling errors also proved helpful when jobs ran into problems. Its proactive e-mail alerting significantly reduced the time involved in identifying and addressing issues.

Apart from specific features, what ultimately influenced the decision to choose Tidal Enterprise Scheduler was UCHC’s overarching need for process automation. Kevin McDonnell, DBA at UCHC, remarks that Tidal Enterprise Scheduler complements elements of the Informatica solution. “Informatica creates staging environments for our data warehouse initiatives. To move from a staging to a production environment successfully means bringing in flat files and cleaning up data. Tidal helps us keep track of all aspects of the process and schedules them in the correct sequence. We’ve largely been able to eliminate manual intervention. That’s a huge bonus.”

Results: Greater Efficiency and Savings across a Complex Organization

Effectively monitoring and scheduling the processes involved in data aggregation and loading ensures that a sophisticated BI solution, such as Informatica, delivers the accurate, timely information that users need.

At UCHC, the IT team has made good use of Tidal Enterprise Scheduler’s automation and standardization capabilities and has extended its deployment use beyond the core user group. In the works are plans to incorporate its help desk in day-to-day Tidal operations.

“The Tidal Enterprise Scheduler has done an excellent job of reducing the amount of time we need to spend in managing our environment,” says Jay Coppola, data warehouse architect, UCHC. “For example—and this is just one saving I can point to—monitoring file size used to take as much as four hours of a team member’s time. Now, Tidal handles that for us. That’s time we can divert to other important tasks.”

Coppola further notes that the scheduler will enable UCHC to work effectively with new applications. “We’ve recently added Oracle to our environment and already use SQL Server. Though both incorporate some native scheduling, we have found that it makes more sense from a management and performance standpoint to leverage Tidal Enterprise Scheduler. We see that as part of our new application-related procedures going forward.”

For a free white paper on this topic, click here and choose the title “Automating Data Flows to Your Business Intelligence System.” For more information about Tidal Software (now a part of Cisco), click here.
Why End-to-End ETL Automation Is a Must for BI Success

By Derek Evan  
Solutions Architect, Cisco

The information retrieved from a BI system is only as trustworthy as the data going into it. Bad information is generally not the result of bad source data; rather, it’s the result of pulling out-of-sync data into the system due to problems in the data processing flow. Process flow issues can also impact information availability, process auditability, and IT’s ability to respond to changing business needs. While bad information can lead to poor decisions, the full scope of problems includes high IT costs, potential governance and compliance issues, and other less-than-desirable consequences.

Scripts, Custom Code, and Islands of Automation

In typical IT environments, there can be hundreds—or thousands—of data sources, including legacy databases, application databases, departmental data marts, and information from business partners. Each can have unique issues around method of access, content, and quality; update and arrival schedules; and requirements for transformation.

ETL tools often include basic process schedulers to initiate, coordinate, and manage the process. These schedulers typically direct operations inside the tool. For process flows outside the tool, scripting and custom code is required. Error handling and documentation are frequently afterthoughts. The overall result is a solution that provides little visibility into the ETL process, scant notification when a portion of the process fails, and no help in recovering from errors.

Recovery and Cascading Errors

Even in a well-managed system, errors and failures inevitably occur. Periodic problems with networks, servers, storage devices, and applications cause processing steps to fail. Data or application errors may produce corrupted data or truncated files. Unplanned maintenance may take a critical application offline. In a system based on scripting, custom code, and embedded schedulers, these numerous problem points make the BI solution so fragile that IT spends significant time identifying and solving problems.

Another major problem is cascading errors. These occur when custom solutions lack sufficiently robust error detection or dependency management to halt execution of a data flow when a step fails. A series of errors in downstream steps then occurs. Some steps fail completely, or even worse, failures go unnoticed and processing continues using erroneous or missing data.

End-to-End Automation Is the Key

Controlling the ETL process flow requires standardizing on a platform that automates and provides visibility into the data flow end to end. A particularly effective tool is a distributed job scheduler. Distributed job schedulers offer both the reach and control necessary to manage the ETL process, as well as the input, output, and notification functions associated with a complete data flow.

Key to proper ETL processing is managing the sequence of steps in the process and coordinating those steps that may operate autonomously. Enterprise job schedulers orchestrate these sequences as a series of dependencies, which can be defined to include a broad range of events and triggers. Through this mechanism, each step in a process is ensured complete and accurate data, and the next step proceeds as soon as possible, maximizing throughput for the BI system. Automated dependency management dramatically reduces the complexity in creating an end-to-end ETL process and is the only way to consistently deliver data to the right place at the right time.

With the automation of an enterprise job scheduler, cascading errors are prevented right from the start. Dependency management ensures that each step in a process completes successfully before launching the next activity. Additionally, errors are isolated quickly, because the scheduling console immediately highlights any failed processing steps, localizing the problem to the specific system component.

For a free white paper on this topic, click here and choose the title “Automating Data Flows to Your Business Intelligence System.” For more information about Tidal Software (now a part of Cisco), click here.
Visionary School District Finds New Ways to Eliminate IT Interoperability Issues

By Richard Krause
Product Marketing Manager, Data Integration Products, Vision Solutions, Inc.

Broward County Public Schools in Florida has developed an innovative student analytics system and data warehouse that captures immense volumes of information, including attendance records, grades, test scores, medical histories, and contact information. More than 200,000 students, partners, teachers, and school administrators have password access to this new data warehouse, where facts are available on demand. To achieve this level of collaborative transparency, Broward needed to eliminate barriers between several disparate databases and platforms.

Replicate1 from Vision Solutions has been implemented at Broward County Public Schools to provide real-time changed data capture and on-demand, enterprisewide data sharing between any application, regardless of database format or operating system.

According to the school district’s senior data warehouse analyst, Phyllis Chasser, Ph.D., “About 85 percent of our users are on Macs, and the other 15 percent on PCs. All of the student transactional applications as well as our data warehousing activities run on IBM System i. With Vision Replicate1, we can bring data over rapidly from a wide range of data sources, including SQL tables, DB2 tables, flat files, SAP, Oracle, and others—often with just a point, click, and drag.”

Vision Solutions helped Broward implement Replicate1. Says Chasser, “We found that a rapid, successful implementation is key to maintaining project momentum, and everything went very smoothly.”

Chasser emphasizes that teachers and administrators can make more collaborative, data-informed decisions thanks to the data warehouse and Replicate1, and this has a very real impact on student achievement.

“A good example is the fact that our district demographics change on a regular basis—we have a lot of students moving from school to school during the year. We used to update all of our records overnight, but now, some of the most essential data, such as student registration and attendance, is real time. So, by the time the student reports to the new school, everything is already there, and if they move during the middle of the day, the records are there immediately,” according to Chasser. “Students can be placed right away in the specific classes they need. If they have problems with English or they need to be on the lunch program, there’s no delay.”

“We used to update all of our records overnight, but now [with data integration], some of the most essential data, such as student registration and attendance, is real time.”

Phyllis Chasser, Ph.D.
Broward County Public Schools

Broward County Public Schools shares their cache of information with an organization called the Broward County Information Network, a consortium of several public entities within the county of Broward, including hospitals. All of these entities access an Oracle database hosted by the county. Chasser explains, “Data sharing definitely makes things easier for agencies across the county. For example, if a child is hurt outside of school, authorities can link into this independent database and get contact information or information for emergency treatment such as allergies that we have in our data warehouse.”

Vision Solutions’ Replicate1 has played a significant role in this evolution. “The data warehouse has created a data culture within the school district, and it’s wonderful. The easy availability of data makes it an integral part of the organization, and Vision Replicate1 has been instrumental in creating that environment of easy availability,” says Chasser.

For a free white paper on this topic, click here and choose the title “Get the Facts: Real-Time, On-Demand Information for Business Intelligence and Data Integration.” For more information about Vision Solutions, Inc., click here.
Advantages of a Software-Based Data Sharing Solution

By Richard Krause
Senior Product Marketing Manager, Vision Solutions, Inc.

Traditional methods of data integration might work for some situations, but could take months to plan and implement. During that time, many changes can occur in your company, your marketplace, and your IT environment, which may impact the scope of the project and lengthen the delivery schedule.

Can your business afford to invest the time and budget resources when speed and flexibility have become competitive advantages? What if you could effortlessly replicate data from legacy DB2 databases into Oracle, Informix, and SQL databases and back again without custom programming or large-scale integration costs? What if you could share data from one application to another in real time and could be assured that the data would be in the exact format the application and users needed?

Would it help if you could make it easier to add a new customer, update an inventory balance, or delete a stock item by having each of those applications share the same information? What would your environment—and the business—gain if you could break down isolated silos of user data and give everyone across the business the same effortless access? What if you could be confident that the right data is reaching the right people so you could spend your time on other pressing IT/business alignment issues?

As executives look to you for cost-effective ways to handle the latest government mandated regulations, data security, and accessibility, would it help if you could meet those demands without costly data quality and integration projects? What if you could accomplish this without discarding your investment in existing systems and databases?

**Why Software-Based Data Sharing/Replication Tools?**

Simplified, on-demand, enterprisewide data sharing with a software-based replication solution fits your needs if you want to quickly solve pressing and ongoing data integration issues between critical applications such as Oracle/PeopleSoft, Siebel, SAP, ERP, and Web and CRM applications. For example, you might need to integrate disparate databases and many critical applications but cannot afford to consolidate them or embark on an expensive application integration project.

Such solutions make data available wherever and whenever it is needed—from and between Oracle, SQL, DB2, Sybase, and other databases simultaneously (and back again if needed). They can solve integration problems between database or application silos that result from recent acquisitions or mergers. They can transform data into usable business information regardless of where it currently resides and which application needs it.

Software-based data sharing/replication solutions can work well in your environment if you want to spend a minimum amount of time setting up and managing the data sharing process. These tools save time and money because they are database driven rather than programming driven. Simplified, on-demand, enterprise data sharing with a software-based replication solution also fully leverages the flexibility of open standards technology into your database management strategies. That can result in fast adaptability to IT changes and future growth plans. It suits companies that need a cost-effective, non-program-specific solution with a fast ROI.

**Break Through Database Barriers and Do More**

Vision Solutions’ Replicate1 replicates data between and among databases in real time or on a schedule using transaction-based or snapshot copy replication. It recognizes data changes when they happen, so it enables real-time replication. Advanced options manage broadcast/cascade replication to conserve network resources, saving you money.

It works with the most popular RDBSs, including Oracle, DB2/UDB, DB2/400, DB2/VMS, Informix, Microsoft SQL Server, and Sybase. This minimizes the need for costly database and application integration. For example, you can set up Replicate1 to transform Oracle data into the exact format needed for DB2 or SQL server databases and applications. Simultaneously, it can transform DB2 and SQL server data into Oracle data types.

For a free white paper on this topic, click here and choose the title “Get the Facts: Real-Time, On-Demand Information for Business Intelligence and Data Integration.” For more information about Vision Solutions, Inc., click here.
The success of loyalty card programs—such as the one instituted by Nectar in the United Kingdom—is highly dependent upon the quality of the data behind them. Through numerous brick-and-mortar and online partners, approximately 50 percent of UK households participate in the Nectar loyalty card program. Since its launch in 2002, Nectar has given back more than £1 billion worth of rewards to collectors. Nineteen Nectar cards are swiped every second of the day.

To join Nectar, shoppers register online or complete paper-based forms available in stores. The Nectar membership database holds details of millions of opted-in households. Nectar uses this information to present targeted, relevant offers to collectors on behalf of the companies that participate in the Nectar program.

“Nectar is trusted by millions of people,” explained Louise Cantrill, campaign services manager at Nectar. “Collector information is our most valuable asset and the quality of that information is everything to us; it’s fundamental to our brand, our partners, and our compliance with the UK Data Protection Act.”

**High Business Value**

By 2004, Nectar was clearly a hit. Households all over the UK were collecting points. Retail and service partners were making as many as 50 extracts a month from their Nectar databases to gain intelligence to guide campaigns.

“With the ever-increasing importance of our collector database to partners, we needed closer control over data quality,” Cantrill said. “We made business management responsible for it and pursued a more systematic approach to spotting issues in records, adding missing information, and identifying duplicates.”

**Automated Data Quality**

Nectar defined, tested, and implemented an extensive and rigorous set of business rules to allow it to better understand the nature of its data and expose and correct issues within it. “Today our quality processes are very advanced, both at a real-time data entry level and on an ongoing batch-maintenance basis,” said Cantrill. “We are confident in our data integrity; quality is now business as usual for us.”

Given the varied quality of source data, Nectar is proud of the accuracy of the contact information it holds. In its active collector database, some 96 percent of its address records are considered to be 100 percent accurate.

**Sophisticated Projects**

Data quality for Nectar may now be a standard process in regard to Nectar registrations and maintaining ongoing data accuracy. But there are also nonstandard projects in which good data quality is essential.

For example, in a multimillion-pound rebrand in 2007, Nectar mailed more than 10 million collectors with new Nectar cards, replacing cards up to five years old. It was one of the largest “recard” mailings in Europe.

For each individual, a personalized card, and later a pack, were produced. These had to then be matched together and mailed. In the intervening weeks, however, there were 300,000 suppressions on the grounds of data latency. Nectar identified these suppressions, saving more than £150,000 in production and mailing costs and ensuring the right cards got to the right people, protecting the integrity of the Nectar brand and reputation.

**Final Word**

The benefits of excellent data quality processes are integral to the success of Nectar. Certainly considerable savings are made through maintaining low rates of returned mail. But much more significantly, Nectar would be seriously undermined if collector data lacked integrity, as this would lead to a loss of faith with collectors and partners.

For a free white paper on this topic, click here and choose the title “The Challenges of Worldwide Data Quality.” For more information about Harte-Hanks Trillium Software, click here.
Demanding Enterprisewide Data Quality

By Len Dubois
VP, Marketing and Sales Support,
Harte-Hanks Trillium Software

Organizations strive to differentiate their products from their competition while offering unique approaches for addressing customer needs. Yet for all companies—large or small, growing or downsizing—there is the same direct correlation between relevant, timely, high-quality data and the company’s ability to succeed in business.

In fact, data assets often become the differentiator for those organizations that compete successfully and increase shareholder value. There is no clearer example of this than in the global financial services market. As we have witnessed over the past year, poor data quality can severely impact an organization, damage a market sector, and deteriorate the public’s faith and trust in the economy.

The types and volumes of data amassed in today’s business systems are varied, numerous, ever-changing, and more complex than ever before. Even sophisticated companies are having difficulty keeping pace with data quality. The use of data and its corresponding informational value changes based on many factors, including the type and number of applications for which it is acquired, accuracy and completeness, and an organization’s ability to share the data across the enterprise.

At the end of the day, though, it is the responsibility of key business stakeholders to ensure the data within multiple, disparate repositories is “fit” for the business purposes for which it was intended. More and more we see business analysts stand side-by-side and collaborate with their IT partners, recommending or implementing changes to the data to ensure alignment with business processes, and always keeping in mind that these changes made to the data at one stage of its lifecycle may impact other processes downstream.

As a result, the marketplace expects a more comprehensive approach to addressing an organization’s data quality needs. An enterprisewide approach to data quality is more closely aligned with the most strategic and operationally oriented initiatives, such as master data management, data governance, and risk management and compliance. It is these types of engagements that enable organizations to move beyond the legacy approach of solving data quality challenges to an approach that instills trust and confidence in the quality of data throughout the organization. This approach also readily derives the most business buy-in for continually improving information quality.

The evolving role of data stewards and business users has given rise to the requirement for data quality vendors to demonstrate how poor data quality manifests across organizations’ strategic operational applications, in addition to the more traditional role of providing the ever-elusive “single view.” Industry leaders must provide a distributed enterprise approach to data quality—one that delivers not only best-in-class technology but also a collaborative environment that enables business analysts to understand the impact of poor data quality through the use of ratings, rankings, and scores tied to key performance indicators (KPIs).

In addition, data governance practices have also emphasized the need and desire to demonstrate the impact of data quality to executive management levels. Data quality providers, whose offerings are tied to data in movement or low-volume departmental views, are simply not equipped to deliver information with the speed, accuracy, and ability to mitigate the risks associated with today’s complex business environment.

The need for high-quality information is only going to continue expanding across the enterprise. Today, organizations require a foundation and platform that allows them to deliver consistent, actionable, and trustworthy information, both for individual data quality projects as well as operational information needs that span the enterprise. To do so, they will need a strategic data quality solution partner that can deliver the people, process, and technology to overcome whatever enterprise data quality challenges an organization may face.

For a free white paper on this topic, click here and choose the title “The Challenges of Worldwide Data Quality.” For more information about Harte-Hanks Trillium Software, click here.
PITT OHIO EXPRESS Delivers Results with Improved Customer Data

Transportation company uses DataFlux dfPower Studio to power customer data integration, turning its improved database into a competitive advantage

Commentary by Justine Russo
Manager of Market Research and Business Intelligence, PITT OHIO EXPRESS

The Business
PITT OHIO EXPRESS is a Pittsburgh-based transportation company that handles less-than-truckload shipping throughout the Mid-Atlantic and Great Lakes states. With 21 terminals in its service area, PITT OHIO makes more than 9,500 daily deliveries to more than 14,000 unique locations.

The Challenge
On any given day, PITT OHIO processes 9,000–11,000 new invoices. Each invoice is entered manually, in a process that has historically given greater precedence to entry speed than to data quality. While understandable from an operational perspective, this rapid-fire data entry approach eventually created a significant duplicate data problem for the company. Each variation in customer name or address information would create a new customer record, meaning that a single entity could be represented by dozens of records in the database system.

With this duplicate data, accurately managing customer information was a critical concern. In an industry where swift turnaround time was crucial, the requirement to dedicate valuable resources to the time-consuming administrative task of managing these customer records had a very real impact on the company’s efficiency.

The DataFlux Solution
PITT OHIO selected DataFlux dfPower Studio to complete its customer data integration initiative. The platform’s intuitive interface allowed PITT OHIO’s business users (those employees whose job performance was tied to the integrity of information) to take responsibility for managing customer data.

DataFlux dfPower Studio allowed PITT OHIO to discover and address problematic data and to verify and merge customer records. Business users built data improvement workflows quickly and logically with the innovative job flow builder.

With dfPower Studio’s fuzzy logic matching capabilities, PITT OHIO was able to reconcile customer records, identify duplicates, and reduce these multiple instances into a single master record.

The Results
PITT OHIO began the project with a goal of a nine-month implementation schedule. With DataFlux’s flexibility and intuitive interface, deployment and learning time were significantly shorter than scheduled. This project, scheduled for nine months, was completed in just five.

“PITT OHIO began the deployment with an expectation of an ongoing 95 percent data consolidation rate,” said Justine Russo, manager of market research and business intelligence. “DfPower Studio has helped us surpass this goal, giving us a consistent 99 percent consolidation rate.” Before implementing DataFlux, PITT OHIO’s customer data management had been consuming significant amounts of time from the IT department and multiple business users. After using dfPower Studio to increase the quality of customer data, PITT OHIO was able to manage this information with only a single business user.

PITT OHIO was able to turn its improved database into a competitive advantage. By refining its address information, the company was able to offer its customers more efficient shipping times and more advanced logistics than its competitors. “Now, with the new, consolidated view of our customer data,” Russo said, “PITT OHIO is able to provide more targeted service, identify our most valuable customers, and deliver competitive value in return.”

Quick Facts about PITT OHIO EXPRESS

• PITT OHIO routinely handles 9,000–11,000 invoices per day
• The company reduced 650,000 unique database entries into an accurate master repository of 10,000 records
• The DataFlux solution helps PITT OHIO maintain an ongoing 99 percent data consolidation rate

For a free white paper on this topic, click here and choose the title “The Three Key Phases for Data Migration—and Beyond.” For more information about DataFlux, click here.
With Data, the Whole Is More than the Sum of Its Parts

Excerpt from The Data Asset: How Smart Companies Govern Their Data for Business Success, a SAS Business Series book by Tony Fisher, forthcoming from John Wiley & Sons in July 2009

Many organizations find they cannot rely on the information that serves as the very foundation of their business. Unreliable data—whether about customers, products, or suppliers—hinders understanding and hurts the bottom line. It seems a rather simple concept: Better data leads to better decisions, which ultimately leads to better business. So why don’t executives take data quality and data governance more seriously?

In my experience, this lack of attention to data severely and negatively impacts numerous organizations. We all need to understand that we are seeing a shift in the way we think about and treat data. Successful organizations are moving from a focus on producing data to a focus on consuming data.

For most organizations, this journey is just beginning. And for many, this journey begins with education built on a solid data foundation.

In organizations today, data is typically somebody else’s problem—until something bad happens. The CEO of a plumbing manufacturer learned this the hard way a few years ago. One of his major manufacturing plants burned to the ground, and he was eager to immediately inform customers of the situation. The CEO asked for a list of products that were expected to be manufactured in the destroyed plant and a list of customers that were expecting delivery.

The finance department provided a list of everybody who had bought something, but that department didn’t know the product delivery schedule. The sales office knew of every customer and what they had purchased, but not where the products would be manufactured. The manufacturing plant had a delivery list of what to produce, but not a full inventory of what was in the production pipeline.

Of course, the closest thing to what the CEO needed—the delivery list—was destroyed in the fire. Eventually, the IT department cobbled together an incomplete list and presented this to the CEO. Predictably, the CEO became frustrated. In the end, he decided data wasn’t such a dull topic after all; it was integral to his business. A quote from the philosopher Aristotle sums up what the CEO—and this entire organization—learned from the experience: “The whole is more than the sum of its parts.” The sum of the data in the individual systems did not accurately depict the whole of the business.

Aristotle was one of the greatest of the ancient Greek philosophers and is still considered one of the most visionary thinkers of all time. As a pioneer in the field of metaphysics, Aristotle sought to develop a method of reasoning that would help in learning as much as possible about an entity.

While most discussions about data do not start with philosophical references, it is important to note that the crux of Aristotle’s philosophy applies to most enterprises. Exhaustive efforts at studying, cataloging, and accessing information led Aristotle to the observation that the whole is more than the sum of its parts. Like Aristotle’s quest to know and understand, data management is about learning everything there is to know about your organization—and more specifically, learning everything there is to know about the data that is required to run your organization.

For a free white paper on this topic, click here and choose the title “The Three Key Phases for Data Migration—and Beyond.” For more information about DataFlux, click here.
The Northern Ontario School of Medicine (NOSM) is mandated to educate doctors and offers an undergraduate MD program, various residency programs, and the Continuing Health Professional Education (CHPE) program for practicing physicians. To manage the CHPE program, NOSM developed a business application to administer information and activities such as course descriptions, participant registrations, and transcript management. However, the system was manually intensive, time consuming, and consisted of a combination of disparate Microsoft SQL Server applications, Microsoft Access data sources, and paper processes. We wanted to improve our program by creating a single application and completely automate the process of scheduling and managing our CHPE events.

Critical Need for Data Quality Processes
To accomplish this application development project, NOSM invited various systems integrators to complete requests for proposals (RFPs) to obtain bids for the project. A critical capability needed was expertise in the area of data quality, as the initial data to be consolidated was not standardized or well monitored.

Ultimately, two systems integrators who work with Melissa Data were awarded the project, in large part because of their expertise in data quality and the strength of the Melissa Data Total Data Quality Integration Toolkit (TDQ-IT) for Microsoft SQL Server Integration Services (SSIS). We found the TDQ-IT components provided the full spectrum of data quality needs within the SSIS data flow, and can simply be added to SSIS as drag-and-drop icons within Visual Studio.

The two integrators, Fox River Software and Actuality Business Intelligence, won the bid based on a proposal to build a custom application using Ironspeed software as the Web site application development framework, Microsoft SQL Server as the centralized database, SSIS as the data integration and workflow tool, and Melissa Data TDQ-IT as the data quality solution.

NOSM was able to integrate data from about 30 distinct source systems.

Specific Data Quality Processes in the NOSM Application
For our application we are using Melissa Data’s TDQ-IT to profile (identify data quality issues), cleanse (test to meet business rules), parse and standardize (restructure data into a common format), match (find unique identifiers and perform deduplication), enrich (phone and e-mail validation), and monitor (check conformance to data quality requirements).

As one example, we utilized the Melissa Data e-mail and fuzzy matching components to take data inputs from six completely different sources (including Microsoft Access, Excel, flat files, and Novell Groupwise), merge them together, validate the e-mail addresses, and create a resulting list of deduplicated e-mail addresses and names. Ultimately, we were able to cleanse and integrate data from approximately 30 distinct source systems, which are now centralized in a new, automated CHPE system, saving large amounts of time and money.

For a free white paper on this topic, click here and choose the title “Six Steps to Managing Data Quality with SQL Server Integration Services.” For more information about Melissa Data, click here.
Finally the Right Tools for the Job: Fix the Flat Tire, Replace the Plugs, or Buy a New Car

By Ira Warren Whiteside
Chief Data Quality Evangelist, Melissa Data

Marketplace
Data quality software, as defined by analysts, consists of data profiling, generalized cleaning, parsing, matching, enrichment, and monitoring. Most clients have implemented each of these processes in some fashion. Some lean more toward custom code, and others toward a mix of vendors. We believe the market is ready for an incremental approach to solving a customer’s data quality problems. The primary goal is to provide the customer with the tools to integrate each entity to the specific level required or needed. Whether customers need a one-off quick fix or a permanent production solution, they should have the ability to use the same tool-set for either problem.

A Conceptual Data Quality Smackdown
All too often we seek an alternative solution instead of repairing or fixing what we already have. From this standpoint, we need more tools and fewer solutions. This may seem counterintuitive in the BI vernacular; however, it is exactly what we do in the real world.

In our current economic market, most clients are interested in fixing their data and leveraging their existing tools with their existing resources—contractors or employees. Let’s look at some real-world examples: You buy a car, it breaks, and you fix it. You build a house, it breaks, and you fix it. Your first option is usually to repair rather than replace. In BI, we take a skewed approach and assume the customer needs a new solution rather than a repair. Perhaps we should look at our customers’ existing applications as their solution (car, house) and concentrate on providing them the means to repair what’s broken.

The fix to a customer’s problem may not be a new solution, but rather a new tool. Most customers have unique requirements and partial solutions in place. By focusing on providing the right tools, customers can begin to repair their existing data quality solutions and gradually improve their processes until they are satisfied. Another issue facing customers in the BI or software world is that they can’t just buy one tool—vendors usually want to sell a complete solution in lieu of a tool or a set of tools. A new approach would be to provide the tools and instructions for customers to help themselves—in essence, a DIY (do it yourself) approach.

In addition, a client may have many platforms in use. If customers take a toolkit approach, they can use the tool in whatever platform they have. For instance, if you need profiling and/or address correction, you want a tool that integrates into all of your platforms (e.g., Microsoft SQL or SSIS, Oracle PL/SQL or OWB, Informatica). It is true that this approach would involve some custom fitting, but just like the real world, if you need a new carburetor or a new water heater, you simply install it.

For a free white paper on this topic, click here and choose the title “Six Steps to Managing Data Quality with SQL Server Integration Services.” For more information about Melissa Data, click here.
New EIM Requirements Demand Better Solutions

By Dennis McLaughlin
VP of Sales, iWay Software
and JT Taylor
CTO, iWay Software

Enterprise information management (EIM) has emerged as an effective way for companies to better control and leverage the explosive volumes of information they generate and maintain. When an organization plans and executes an EIM strategy properly, it can significantly enhance the value of its corporate data, creating a “single version of the truth” for all who access and utilize information, company wide.

However, many organizations embarking on an EIM initiative have a narrow view of what it really is and often have selected and deployed solutions that address only one piece of the puzzle. Their plans—and the technologies they have acquired to support them—have gaps that make the achievement of true, enterprise-scale information management an elusive goal.

Every Piece of Information Counts
Many companies build data warehouses and marts to serve as the foundation of their EIM strategies—but do little else to facilitate the efficient creation, maintenance, and use of corporate data. This approach has been proven ineffective because these sources have been populated with only historical data from back-end systems.

What the approach ignores is the critical data that constantly flows across various points during the course of business activities (e.g., information generated during automated workflow execution or consolidated data available for decision support). It also fails to consider the data that enters and moves throughout the business via electronic means, such as EDI and other business-to-business (B2B) transactions.

Additionally, EIM must not only address structured information stored in databases, transactional systems, marts, and warehouses, but also information contained in spreadsheets, EDI documents such as sales orders, and other semi-structured data. In order for EIM to have maximum impact, all data must be managed from end to end, regardless of its point of origin, source, format, or location.

Quality Management Must Be Proactive
Any EIM strategy that focuses solely on the collection, consolidation, and centralization of data will be only partially successful and deliver moderate returns at best. Data must be more than just accessible. It must be consistent, accurate, and relevant—requirements have become increasingly urgent as reporting and information disclosure guidelines continue to grow more rigid. Companies must also guarantee that corporate data is always correct, complete, and most important, auditable.

But as information passes through a larger number of touch points (e.g., as it moves among more systems or more people interact with it daily), optimum integrity and consistency become much harder to achieve and maintain.

What many companies don’t realize is that once corrupted, outdated, or invalid information enters their environment, the potential damage may have already been done. For example, bad information may have been used to make a critical business decision. It may have resulted in a missed opportunity, or even worse, it may have been shared with regulatory agencies, hindering compliance and putting the company in jeopardy of receiving fines or penalties.
Data must be managed proactively, not reactively. This means bad data must be identified and corrected before it finds its way into an infrastructure.

Additionally, the majority of data quality solutions are applied only to the information contained within warehouses and other sources designed for end-user access and decision-making support. But information contained in back-end systems also plays a crucial role in core business operations. Therefore, data quality must be managed at the point of origin—not just the point of access.

“Real Time” Is Critical

Many people associate data warehouses and marts with historical information. While this may work for some needs, there are many scenarios in which information needs to be managed in a true real-time fashion. For example, information derived from certain B2B transactions, such as sales orders, may need to be dynamically pushed to suppliers or pulled by production automation systems the moment transmission takes place.

Since information contained in warehouses is rarely current, the additional burden is placed on back-end systems, negatively impacting their performance. However, when an EIM strategy supports real-time information capture and handling, data is instantly available to facilitate core operations, and the strain on back-end databases and data warehouses is eased, freeing them up for their primary purposes—transaction processing and decision support, respectively.

What You Need in an EIM Framework

Because traditional EIM requirements are changing drastically, the solutions employed to satisfy them must be improved and more comprehensive. To address these new needs, today's EIM suites must provide such advanced functionality as:

Unlimited Information Access

Companies own countless types of information assets, including packaged applications, databases, files and documents, electronic messages such as SWIFT and HIPAA, and automated transactions. An EIM framework can only deliver significant results if it can handle all the types of data that exist across a business, regardless of who generated the data, where it's stored, or whether or not it's in a standardized format.

Support for Multiple Levels of Information Latency

A world-class EIM suite will be highly flexible, allowing companies to manage all data and its quality via a variety of methods: online as users are generating it or interacting with it, in real time as it is created during automated processes, or in batch as it is pushed or pulled across various sources at timed intervals.

Proactive Data Quality Management

A variety of data quality techniques, such as profiling, automatic cleansing, and dynamic merging and matching must be utilized to not only locate and remove bad data from the environment, but also to prevent it from entering in the first place. Only this kind of truly proactive quality management will ensure optimum integrity of enterprise information.

Additionally, an EIM framework should provide the ability to enrich and enhance information. For example, it should allow for the extension of existing corporate data with additional relevant information derived from third-party sources.

Master Data Management

An EIM suite must enable the creation of a single system of record—whether it’s a central physical master data instance or a “virtual” repository—that can feed complete, consistent, and correct data back to applications across the company. This is particularly important in certain key functions, such as customer relationship management, financial management, or inventory management, where numerous disparate systems, maintained by different departments, may contain multiple versions of the truth that can negatively impact related business operations.

Flexibility and Reusability

EIM is a dynamic concept and its requirements continue to change drastically. What EIM means today will likely be radically different from the EIM of the future. Therefore, any EIM suite implemented to support current needs must be powerful and agile enough to address new needs as they emerge. Additionally, an EIM suite that is built on reusable components will deliver the greatest current value and maximize long-term return on investment.

Summary

In order to achieve EIM success, the strategy—and the technologies that support it—must account for all types of information, all levels of data latency, and all middleware paradigms. Additionally, to effectively create, administer, and use information on an enterprise scale, both data quality and master data must be thoroughly managed from end to end. Only by addressing these needs will an organization realize true information consistency, accuracy, and accessibility across their entire business.

For a free white paper on this topic, click here and choose the title “Enterprise Information Management (EIM): The Hidden Secret to Peak Business Performance.” For more information about iWay Software, click here.
Selling MDM to the C-Suite

By Jim Walker
Senior Marketing Manager, Initiate Systems, Inc.

Company Overview
A diversified financial services company offering a broad array of credit, savings, and loan products to customers in the U.S., UK, and Canada.

The Challenge
This company faced a significant challenge in substantiating the business case for MDM. However, using some best practices for selling the project, they successfully acquired funding and increased visibility within the executive level of the company. This case study outlines their approach and provides a basic game plan for getting your MDM project sold in the C-suite.

The Solution
There are a few key concepts that helped the MDM project leaders build and sell their successful MDM business plan. First, they demonstrated the value of their project in business terms and mapped this to corporate priorities. Second, the plan provided a clear definition of short-term value while painting a picture for ongoing, incremental value over time.

Mapping Projects to Priorities
By mapping their MDM project to corporate objectives, the project leaders were able to demonstrate expected value that executives could categorize and understand.

To get started, this organization gathered a cross-functional group of both IT and business owners to gain a common understanding of MDM and then outline the impact for their respective groups. The result of this meeting was a “greatest hits” of MDM projects categorized under corporate priorities:
Priority: Increase Customer Satisfaction and Retention
• Increase products per customer through improved marketing of cross-sell, up-sell offers

• Consolidate CRM activity and data across lines of business (LOBs)
• Provide consolidated customer account statements and accurate, consistent customer information through all channels and at all touch points
• Create the “right” products for the “right” customers
Priority: Reduce Costs
• Consolidate service centers and reduce customer service wait times by providing consolidated customer information to sales representatives
• Reduce marketing costs by avoiding redundant mailings
• Consolidate third-party data, eliminating overlap and saving license costs

Priority: Identify and Mitigate Risk
• Derive accurate customer profitability and risk profiles
• Implement cross-account and LOB collections and recoveries

Priority: Comply with Regulations
• Maintain a single source for customer privacy preferences across systems and LOBs (opt in/opt out)
• Comply with anti-money laundering and “know your customer” regulations and initiatives

Constructing the Plan
Once the list was compiled, the next step was to identify a set of projects that would demonstrate the greatest value and could be implemented quickly. This “short time to value” list was accomplished by compiling three identified projects across three different corporate priorities. The phase one business plan contained these three objectives:
• Consolidate CRM activity and data across LOBs
• Derive accurate customer profitability and risk profiles
• Manage customer privacy preferences centrally

Short-Term Pain, Long-Term Gain
This comprehensible set of projects/values outlined a quick time to value. The MDM team augmented this section of the business plan with a list of all potential projects and outlined how the first phase project institutes an MDM foundation that additional projects can be built upon in the future.

MDM Game Plan
BUILDING A BUSINESS CASE

1. IT and business owners collaborate to brainstorm MDM projects that will benefit the organization. Assign value and length of implementation time for each.

2. Prioritize collected project list and map each proposed project to the corporate objectives.

3. Choose a set of projects that addresses several corporate objectives and will show valuable return in the short term.

4. Build and sell the plan, which should demonstrate resolution of short-term pain while setting up a plan for long-term value derived by the MDM platform.

The team presented the complete “greatest hits” list, assigning each to a subsequent phase so the executives could see future value. This allowed the MDM team to:
• Demonstrate incremental value over time
• Allow the initial business case to fund an enterprisewide MDM infrastructure
• Help executives envision value and short-term gains, while understanding long-term goals
• Provide smaller, detailed business cases with the projects people can rally around
• Facilitate commitment to dates and set clear, obtainable milestones

Conclusion
After a successful phase one, the customer has started to provide value on phase two projects. An MDM infrastructure using service-oriented architecture (SOA) is in place and the project has transformed their business processes. Selling this strategic asset to executives is a challenge, but with some careful planning and consideration, it can be successful.

For a free white paper on this topic, click here and choose the title “Master Data Management as ‘Plan B’: Why Your Data Warehouse, CRM, ETL, and EII Solutions Are Better with MDM.” For more information about Initiate Systems, Inc., click here.
Multidomain Master Data Management

By Jim Walker
Senior Marketing Manager, Initiate Systems, Inc.

Mastering multiple domains is the natural extension of a successful first phase of an MDM project. Typically, organizations start mastering a single domain, such as a person. Once successful, the organization quickly realizes that the addition of location and products or services can solve complex business issues and also open new models and opportunities to extend value across the enterprise.

The value of multidomain MDM is derived from the ability to extract insight from the complex relationships between customers, parties, external suppliers, partners, and the complexity of the goods and services they provide.

This insight will provide substantial business benefit to the organization. First, mastering a domain provides the enterprise with a single version of the truth for the entity, resulting in reduced duplicates, fewer inconsistencies, and an accurate, complete representation of important data. Second, the relationships between these mastered domains can support and improve valuable operational and analytical business processes, resulting in tangible improvement to the bottom and top line.

Organizations are starting down the multidomain path to reap these benefits today. For instance:

- A service organization has optimized operations and improved service by using location to master multiple customers under a single roof.
- A criminal justice agency is protecting its community by connecting the dots between people, places, things (such as weapons or vehicles), and events to more effectively identify and catch criminals and persons of interest in real time.
- A retail business is reducing risk, improving sales effectiveness, and increasing customer satisfaction by using customer, product, and location information to gain a reliable understanding of products and services available at a location. The business then creates marketing programs to help pitch the “right” product at the “right” location for the “right” customer.
- Healthcare enterprises are able to properly identify providers and their relationships with patients and to associate providers with outcomes to measure the value of the care provided.

Organizations that decide to master more than one data domain are best served by deploying an MDM architecture that has a coherent, unified approach to dealing with all master data types. Mastering multiple domains requires an MDM solution that provides the following:

- **Flexible data model.** As you add domains to your implementation, the model must be flexible enough to accommodate whatever data you want to master. Additionally, the solution should provide standard data model templates as a baseline for quick delivery time and allow you to easily modify the templates to suit your particular data needs.
- **Open integration options.** Every organization is different, and interfacing with various data sources is complex business. Your chosen MDM platform should offer multiple production-ready integration options, from customizable Web services to standard ETL and batch-based interfaces.
- **Accurate relationship engine.** Your solutions’ matching algorithm should be sophisticated, yet easily tuned so it can automate match and link functions as accurately as possible, no matter the data model. This not only helps your business identify unique customers and products, but it also eases the burden of data stewardship.
- **Scalable, high-performance architecture.** Adding domains and data requires analytic horsepower. The solution should scale to meet the needs of a growing organization. It should support bulk data migration as well as sub-second response times over hundreds or even billions of records.

Mastering customer data is the logical first step down the MDM path for many organizations. Extending MDM to cover both people and the relationships between people, places, and things is a valuable step in the right direction for your MDM journey.

For a free white paper on this topic, click here and choose the title “Master Data Management as ‘Plan B’: Why Your Data Warehouse, CRM, ETL, and EII Solutions Are Better with MDM.” For more information about Initiate Systems, Inc., click here.
MDM Brings Financial Services Closer to Customers

By Baseline Consulting

As one of the world’s largest credit unions, LCU boasts close to half a million customers (“members” in its vernacular) and has more than 50 branches across its region. For the past several years, the financial services industry has had a laser focus on customer centricity. At LCU, however, customer information was scattered throughout business systems that collectively did not recognize the customer as unique across the enterprise.

Numerous interviews with LCU managers revealed that customer information:

- Was not always available in the time frames required
- Had data accuracy and quality issues
- Was inconvenient or difficult to access, causing some staff to stop looking
- Was not always presented in a way that supported customer service needs in real time

Clearly, there was a need to provide information to LCU’s staff to help obtain a more unified perspective of the member—that is, across all of the member’s interactions with the credit union. This unified view had to support both operational and back-office processes, and focus on a more analytical review of customer information for improved reporting, advising, and “after the fact” decision making.

Baseline Consulting was engaged to assess LCU’s challenges and recommend an integrated solution. More than 30 interviews were conducted with all levels of management. Consistently, we heard that the key to ensuring a member-centric focus was to provide LCU’s member-facing employees with the information to make decisions “in the moment.”

Baseline determined that current systems could not be cost-effectively modified to support the desired member-centric approach, and thus proposed a customer master data management (MDM) solution implemented in a service-oriented architecture (SOA) environment. By integrating, merging, and deduplicating all of the customer information from the functionally stove-piped business systems, the MDM hub would provide the unified view of the customer that would not otherwise be possible. (Both MDM and SOA were completely new technology initiatives for the credit union.)

Combined with a new portal-based, front-end application, the resulting solution—dubbed MemberView—provided the unified view of the customer that had been missing. MemberView offered an integrated working environment that:

- Provided an immediate display of accurate, consistent, and relevant member information when needed (e.g., member demographics, products and services, relationships with other members)
- Allowed the employee to jump to the right application based on the task at hand, transparently providing the necessary member “context” in the legacy system without causing the employee to search again
- Implemented single sign-on, which improved employee productivity by eliminating the necessity to explicitly sign on to multiple business applications

The consolidated, deduplicated customer information in the MDM hub was integrated, via workflow, with existing legacy systems and the data warehouse, and even more benefits were recognized: improved change-of-address processing across systems, better information for credit risk assessment, and improved regulatory compliance.

The business case used to secure funding for the project highlighted the most important benefit of this solution: “MemberView will help to transform us to a new paradigm where staff will focus on building upon relationships with its members and with the communities they live in, in contrast to helping members perform transactions on their accounts.”

MemberView also improved the member experience in a number of different ways:

- Higher levels of service because the staff had a more complete picture of members’ relationships to LCU across all lines of business

- Faster transaction service in the branch, which increased the time available for advice on more effective utilization of LCU’s products and services

- The right marketing offer for products that were appropriate to members, in ways chosen by the member (e.g., members did not receive offers about products they already owned)

- More comprehensive advice that met both retail and business needs

- A consistent service experience across retail and business channels

Since MemberView’s benefits extended to most of LCU’s staff and operations, it allowed for other indirect process improvements that were previously not possible. For instance, as a member-owned credit union, LCU must conduct regular elections of their board of directors. This election process—which was previously conducted via mail, but had shifted to online voting as the primary mechanism—is critically dependent on a consolidated, deduplicated list of unique members. This ensures that every eligible member gets to vote once, and that no ineligible members or non-members are allowed to vote. Prior to MemberView, assembling and deduplicating the list of eligible voting members was a highly error-prone process. That process has now been streamlined.

For a free white paper on this topic, click here and choose the title “Managing Data, Managing Change: Making MDM Work.” For more information about Baseline Consulting, click here.
Righteous Data:
The Business Value of Data Integration

By Jill Dyché
Partner and Cofounder, Baseline Consulting

Every Sunday when I was a kid, my parents dropped me off at church. My dad would pull up, deposit me in front of the wide wooden doors, and drive away toward errands unknown. This weekend ritual endured until I went to college, where I promptly stayed. I sinned often and I sinned well. My parents were stunned. They’d not only sent me to church every Sunday, but also shipped me off to church camp for two weeks every summer!

It seems as if no one has anything nice to say about their corporate data these days. They don’t do much with it, assuming they can even get it. We’ve spent so much money and time on our analytic platforms, but our data still disappoints on many levels. To echo my parents’ lament, “Where did we go wrong?”

One problem is that until recently, corporate executives have only been paying lip service to data. Sure, they fund the platforms and may even give us some tools. But they haven’t treated corporate data as an asset in its own right, worthy of discrete investments and skills. When it comes to data, they want everything to work out automatically. They want to send it off to church and have it return righteous.

Nowadays, managers in both business and IT organizations are realizing that they not only have to believe in a higher purpose for corporate information, but they also have to change their behaviors—and they are. As we learned together with our credit union client, the pressures calling for better, more useful, and more integrated data are bearing down on businesses. We’ve seen four main business forces that are punctuating the need for more integrated data:

1. Executives rethink profitability measures. Managers are, often for the first time, calling into question the assumptions that have heretofore informed their decision making around corporate profitability. Lofty, long-term revenue goals are quickly ceding to a focus on short-term profit. This means getting tactical. Companies are investing in fixing their data, not only to support immediate business actions such as corporate acquisitions and restructuring, but also to prepare themselves for the next wave of growth. As Jack Welch famously said, “When times are good, you grow. When times are bad, you build.”

2. CIOs search for efficiencies. A few years ago, a study estimated that more than 60 percent of systems development effort was spent gathering and integrating data. Multiply that by the number of concurrent development efforts in your IT department and you’ve got a lot of people duplicating effort—and data. CIOs are looking at master data management (MDM) to mitigate redundant tasks and reduce development costs—and to reconcile heterogeneous data. As MDM offers a single version of the truth to a range of applications across the enterprise, it reduces development time and allows IT managers to use their resources more effectively.

3. Management reconsiders core competencies. “Keep core activities in house, and send everything else to Bangalore,” is a well-worn outsourcing adage. It seems like some managers only parsed the last half of that sentence, not stopping to consider what their core activities should be. Now they’re starting to ask, and they’re modifying their policies. “Keep the data in house, and outsource commodity work around it” is an increasingly popular philosophy that not only promises economies of scale, but also budding information expertise across both business and IT.

4. The business gets impatient. Mergers and acquisitions. Customer retention. Cross-selling and up-selling. Regulatory compliance. The need to synchronize information with strategic business processes is leading savvy managers straight to MDM. Companies have practically codified the “know your customer” rule and have put the data governance and integration processes in place to ensure its adherence. The whole being greater than the sum of its parts, MDM hubs, service orientation, and data profiling and quality tools are (finally!) getting the attention they deserve.

As business professionals become more adept at decision making, they know their data better and can govern it—implementing the oversight and policy-making processes that support enterprisewide information. They can manage, correct, and enrich that data so that it’s consistently more valuable over time. They can even decide to leave that data to its own devices and see how it fares, perhaps realizing in hindsight where they went wrong.

For a free white paper on this topic, click here and choose the title “Managing Data, Managing Change: Making MDM Work.” For more information about Baseline Consulting, click here.
Compliance and Risk Management: Ten Critical MDM Capabilities to Include in Your Evaluation

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

The financial sector meltdown of 2008 and subsequent worldwide economic downturn has organizations across all business sectors bracing for tighter regulations and increased compliance demands. It raises a question for compliance officers and senior decision makers: Exactly what kinds of actions are legislators and government regulators likely to take? New agencies? Broad new legislative actions? Smart decision makers, irrespective of their industry, are seeking technology investments to help establish good governance models that in turn will help with regulatory compliance and lower their operational risk, no matter what kind of new regulatory environment emerges.

Effective regulatory compliance monitoring and reporting requires a strong combination of people, processes, and technology. But in today’s large organizations, the key to maintaining compliance is the proper management of data. Organizations with a strong data governance regime and the technological capability to carry out compliance initiatives are the most successful in meeting regulatory requirements. Master data management (MDM) is a valuable technology for firms in regulated industries because it provides both the framework for effective data governance and the technological underpinnings for compliance operations and reporting. MDM ensures that critical enterprise data is validated as correct, consistent, and complete when it is circulated for consumption by internal or external business processes, applications, or users. But not all MDM technologies are capable of addressing the various compliance requirements facing businesses today. Only an integrated, model-driven, and flexible MDM platform that is easily configurable can provide the functionality required to meet compliance requirements and lower risk.

**MDM Technology = Master Data Governance**

The critical functionality of MDM can be overlooked when companies focus too narrowly on near-term requirements within a single compliance endeavor or business data type, such as customer (customer data integration) or product (product information management). Firms that do so run the risk of investing in technologies that are not easy to extend to other compliance-related efforts across the organization. To reduce the risk of choosing the wrong solution, it is important for firms to consider key business data requirements across several key business functions, including sales, marketing, customer support, and compliance.

To avoid the common mistakes made by MDM software evaluation teams and ensure long-term success, you should make sure that key components are built into your master data management solution. By considering these 10 critical MDM capabilities in your MDM platform evaluation, you will be well on your way toward laying the foundation for a complete and flexible MDM platform that not only addresses your current requirements, but also can evolve to address unforeseen future data integration requirements.

1. **Manages multiple business data entities within a single MDM platform.** Using an MDM platform that can handle multiple data types, an organization can begin to ensure compliance within a single business division to demonstrate a rapid return on investment and later extend the solution to accommodate other business divisions for even greater enterprise value.

2. **Permits data governance at both the project and/or enterprise level.** It is critical that the underlying MDM platform is able to support the compliance-related data governance policies and processes defined by your organization.

3. **Works with your standard workflow tool.** Workflow is an important component of both MDM and data governance because it can monitor compliance in real time and automatically alert the appropriate personnel of any potential violations.

4. **Handles complex relationships and hierarchies.** Certain compliance initiatives, such as Basel II, require the ability to manage complex legal counterparty hierarchies. Make sure your MDM solution is capable of modeling complex business-to-business (B2B) and business-to-consumer (B2C) hierarchies within the same MDM platform.

5. **Provides support for service-oriented architecture (SOA) services.** Since MDM is the foundation technology for providing reliable data, any changes made to the MDM environment will ultimately result in changes to the dependent SOA services, and consequently to the SOA applications. You need to ensure the MDM platform can automatically generate changes to the SOA services whenever its...
data model is updated with new attributes, entities, or sources. This key requirement will protect the higher-level compliance applications from any changes made to the underlying MDM system.

6. **Allows for data to be cleansed inside the MDM platform.** Data cleansing needs to be centralized within the MDM system to provide clean data for compliance reporting. If your company has already standardized on a cleansing tool, then it is vital to ensure the MDM solution provides out-of-the-box integration with it in order to leverage your existing investments.

7. **Enables both deterministic and probabilistic matching.** To achieve the most reliable and consolidated view of master data for compliance purposes, the MDM platform should support a combination of these matching techniques, with each being able to address a particular class of data matching. A single technique, such as probabilistic, will probably not be able to find all valid match candidates, or worse, may generate false matches.

8. **Creates a golden master record with the best field-level information and stores it centrally.** It is important that the MDM system is able to automatically create a golden record for any master data type (e.g., customer, product, asset) to enable compliance monitoring and reporting. In addition, the MDM system should provide a robust unmerge functionality to roll back any manual errors or exceptions.

9. **Stores history and lineage.** The ability to store history of all changes and the lineage of how the duplicate has merged is a necessary requirement to support compliance. Any compliance initiative will require the ability to audit such data changes over several years.

10. **Supports both analytical and operational usage.** Compliance monitoring is performed within an operational system, while compliance reporting is performed using a business intelligence tool or data warehouse.

**Successful Regulatory Compliance Begins with an Integrated and Flexible MDM Platform**

Taking the time to build the foundation for a sound master data governance program is critical to the success of any compliance effort. These 10 requirements will enable you to identify and evaluate a suitable MDM technology platform—a prerequisite when managing your organization’s master data assets and establishing a consistent master data foundation.

Once your organization starts to make its departmental compliance projects operational, you are likely to find that your larger compliance requirements will expand to include other lines of business or geographies. Therefore, it is crucial to carefully evaluate the MDM options and choose a solution that will include all 10 critical requirements.

It is also important to assess the MDM platform’s ability to support these 10 core capabilities out of the box, as they should be integrated components of a complete enterprise wide MDM platform. In this way, you will be able to mitigate technology risk and improve your return on investment since additional integration and customization will not be necessary to make the system operational.

Another benefit gained by having these 10 MDM components integrated within the same MDM platform is that software deployment is much faster and easier to migrate over time. Finally, it is wise to check vendor references to evaluate the enterprisewide deployments of their customers, and to ensure that the vendor’s MDM solution is proven and includes all 10 enterprise MDM platform capabilities.

For a free white paper on this topic, click here and choose the title “Seven Ways to Reduce IT Costs with Master Data Management.” For more information about Siperian, Inc., click here.
Prime Numbers Technology
Revolutionizes the Analysis of Corporate Travel Spending with Open Source Data Integration

Open source data integration simplifies, streamlines, and improves business travel programs

Commentary by Rock Blanco
President, Prime Numbers Technology

A Revolutionary Approach to Corporate Travel
Prime Numbers Technology is among the world’s largest global benchmarking tracking services. The company’s flagship product, Travel GPA, offers a complete travel program management and analysis service that provides in-depth analysis of a company’s overall travel program performance based on policy compliance and performance against other companies in the Travel GPA database. Overall, the system tracks all trip data, customer activity, and customer profiles to help agents make better decisions faster.

Real-Time Travel Data
In today’s fiercely competitive travel industry, margins are tighter than ever and unforeseen events unfold at a phenomenal rate. Travel GPA leverages real-time travel data and history from tens of thousands of companies around the world and provides a basis for establishing travel policies and programs that can improve a company’s bottom line. The data can be scored against internal goals or benchmarked against the Travel GPA database, which comprises more than 17,000 companies and totals more than $4 billion in annual travel spending.

Travel GPA Runs on Talend
After years of custom coding, Prime Numbers Technology needed to find a more robust way of developing and integrating their data integration processes. The company evaluated Talend’s open source data integration solutions based on recommendations from peers in the travel industry. After a few weeks of prototyping, Prime Numbers Technology invested in a subscription license of Talend Integration Suite—Talend’s enterprise-grade data integration platform. Talend Integration Suite extends Talend Open Studio with value-added features for job deployment, scheduling and monitoring, a multi-user development repository, technical support with guaranteed service-level agreements, and IP indemnification. “The fact that Talend Open Studio and Talend Integration Suite share the same development environment made it completely seamless to move from one to the other, and we were able to deploy immediately all the prototyping work we had done with the free product,” said Rock Blanco, president of Prime Numbers Technology.

Talend Integration Suite is now central to Travel GPA, automatically consolidating data from multiple feeds. “We’re handling billions of transactions,” said Blanco, “and we interface with a huge number of diverse sources and a variety of data. We might be dealing with 45 different currencies and perhaps 18 different languages or we might need to merge an HR feed to a particular system. Before Talend, and using the tools that were available at the time, it took us 3 to 4 years to build the 15-plus different interfaces we needed. We replicated all those interfaces on Talend in three months. That’s more than a 90 percent increase in productivity!”

“Talend pulled the travel industry out of the Middle Ages.”

Rock Blanco
President, Prime Numbers Technology

Speed and the ability to reuse code weren’t the only criteria. “Our database is growing exponentially because every second someone, somewhere in the world, is sending data to Travel GPA, and Talend Integration Suite keeps it constantly updated,” Blanco added. “In terms of data mining, Talend pulled the travel industry out of the Middle Ages.

“Our business is data,” he continued. “With Talend we’ve got a fast, reliable dashboard that gives companies a consolidated view of their performance in real time. Using Talend’s connectors I can take extracts from any source.

“Talend has given us the ability to reach into markets we couldn’t target before,” Blanco concluded. “We didn’t have the right tools to handle data collection and consolidation in an appropriate time frame with the resources we had available. We would have needed an entire department and a huge budget to do what Talend Integration Suite lets us do with just a handful of people. Talend is the reason that we’re in business today.”

For a free white paper on this topic, click here and choose the title “The Role of Open Source Data Integration, An Analyst White Paper.” For more information about Talend, click here.
Redefining the Data Warehouse Appliance

By Philip Russom, Senior Manager, TDWI Research

The data warehouse appliance (DWA) has undergone considerable evolution since first appearing early this decade. The evolution is driven by both vendor products and user best practices, and the resulting change has affected the capabilities of DWAs, the type of vendor that provides them, how to use DWAs, and how to define them. Before we list the diverse product types that a new definition of data warehouse appliance should encompass, let’s examine some of the trends and developments that are driving a redefinition.
Whole-Technology Stack Appliances
Netezza was the first vendor to offer a data warehouse appliance (introduced around 2002), so early DWA definitions were based upon Netezza products, which provide a whole-technology stack for data warehousing. That is, the Netezza Performance Server combines database and operating system software with server and storage hardware in a complete data warehouse platform. Before Netezza, Teradata, Sequent, and White Cross (now Kognitio) had for years offered similar single-vendor combinations of hardware and software purpose-built for data warehousing, though not necessarily in an appliance package or described as an appliance.

DATAAllegro launched in mid-2005 with a whole-technology stack solution involving proprietary hardware, similar to Netezza. Although DATAAllegro soon left its own proprietary hardware in favor of commodity hardware from other vendors, DATAAllegro’s DWAs still require specific certified hardware configurations that constitute a whole-technology stack. Microsoft acquired DATAAllegro in 2008 and announced it will fold DATAAllegro’s MPP architecture into SQL Server, which runs on commodity hardware.

Kognitio also jettisoned its proprietary hardware in a process similar to that of DATAAllegro, and Teradata just announced a new family of data warehouse packages that includes DWAs, some running on commodity hardware. Examples of commodity hardware include general-purpose servers built around Intel or AMD CPUs, and popular network and storage hardware from Cisco and EMC.

The movement from proprietary to commodity hardware has good reasons behind it. Commodity hardware is relatively inexpensive and thus helps keep down the price of DWAs. This is important since DWAs compete largely on their low cost. Also, the vendors providing commodity hardware have proved to be good partners for software-oriented appliance vendors.

Partial-Technology Stack Appliances
Starting in 2006, a new wave of vendors emerged with database management systems (DBMSs) purpose-built for data warehousing. These include DBMSs based on the relational model (Greenplum and Kognitio) and the columnar model (ParAccel and Vertica). Most of these DBMSs are licensed in multiple ways. For instance, ParAccel offers three types of licenses:
- Sold and licensed stand-alone, like any DBMS software
- Added onto another database product as a query accelerator
- Embedded in an appliance, usually with a certified or recommended hardware configuration

In the context of the embedded license model, most of the new DBMS vendors call their product a software appliance. This somewhat oxymoronic term refers to a software component (namely a DBMS) that may be embedded in a full data warehouse appliance. Hence, each of these vendors offers a partial stack appliance, often called a software appliance.

The software appliance has proved to be a good starting point for the new DBMS vendors. It allows them to focus on database software (not designing and building hardware), which is their point of greatest innovation and therefore their value proposition. The software appliance product enables the new, small DBMS vendors to partner with commodity hardware vendors and to benefit from these larger firms’ resources.

Miscellaneous Appliances for Data Warehousing or Business Intelligence
A couple of large BI vendors have rolled out specialized appliances in the last few years. For example, Cognos Now! (acquired from Celequest) is a server blade that includes a 64-bit in-memory database for real-time operational BI and performance management with IBM’s Cognos BI platform. It also includes a tool for designing dashboards and similar applications. SAP Netweaver Business Warehouse Accelerator is a similar blade product that accelerates the query performance of the data warehouse component (often called BW) that’s part of SAP Netweaver.

The Sartori Server from Dataupia is hard to categorize. It’s an appliance in the sense that it folds hardware and software components into a rack-mount server blade. It’s an assemblage of commodity hardware parts, such as CPUs, hard drives, and RAID controllers, plus Dataupia’s embedded DBMS. However, it doesn’t stand alone like most DWAs, and it’s usually added on as a query accelerator or capacity extender for other traditional technology stacks.

Hardware/Software Bundles that Resemble Appliances
As DWAs entered the marketplace, relational database vendors (IBM, Microsoft, Oracle, Sybase, and HP) stepped up their offerings of hardware and software bundles that assemble a whole-technology stack for data warehousing. Examples of bundles from leading database vendors include HP Neoview and IBM InfoSphere Balanced Warehouse (most hardware and software components in these bundles are HP and IBM products, respectively). Launched
in late 2008, the HP Oracle Database Machine and the HP Oracle Exadata Storage Server are both based on hardware from HP and software from Oracle. Sybase’s Analytic Appliance (announced in May 2008) combines pSeries hardware from IBM with Sybase IQ (the “mother” of all columnar databases, purpose-built for data warehousing).

Most of these bundles are not DWAs per se because their components are rarely purpose-built for data warehousing, yet they offer many of the benefits of a DWA. In particular, a preconfigured technology stack reduces system integration work, reduces time to use, and comes from a single vendor, which supports the whole stack. Furthermore, vendor size matters in that some user organizations avoid start-up vendors. For these users, the hardware/software bundles are significant, because they come from large, stable vendors and include familiar, mature DBMSs.

**User Practices for Data Warehouse Appliances**

Based on user interviews and vendor briefings, TDWI Research knows there are prominent “sweet spots”—application situations in which users turn to data warehouse appliances:

**Most appliances support a multi-terabyte data mart.** The mart enables an analytic application typically focused on high-volume analysis of customers, transactions, call-level details, utility grid capacity, etc.

**Users begin with multiple terabytes, instead of building up to them.** In the early days of DWAs, most users rolled out 1–3 TB in the first phase of system deployment and grew toward 10 TB. Today, 10 TB or more in the first phase is common, and DWAs can handle this.

**DWAs are used for highly dynamic data analysis, and seldom for reporting.** Typically, a handful of business analysts and similar users are on a discovery mission in which they think up ad hoc queries and alter them iteratively. DWAs are known for perky responses to such complex analytic queries against large data sets.

**Users take full advantage of DWAs’ MPP shared-nothing architecture.** After all, this is what enables DWAs to manage multi-terabyte databases and very complex queries against them.

**DWAs complement and augment enterprise data warehouses (EDWs).** Some appliances host an EDW, so that’s possible. But the vast majority of DWAs are deployed as an SOS—a “system on the side” that off-loads data management tasks and analytic workloads that are best kept out of the core EDW. Hence, the DWA has joined other SOS platforms in the extended EDW environment, such as operational data stores, data staging areas, data marts, cubes, and so on.

**Some DWA users follow a load, analyze, and delete (LAD) method.** For example, when a business problem or opportunity arises, business analysts extract terabytes of relevant operational data and load it into the DWA. They then analyze the information until they learn what they need to know. Before moving on to the next analytic project, they delete (or archive) the terabytes they’ve been working with and start over with a new multi-terabyte data set.

**DWA-based analyses work with operational data in poor condition.** Note that the LAD method doesn’t allow time for much data modeling, transformation, or cleansing. Luckily, DWAs compensate for less-than-ideal data structures and quality by supporting highly complex SQL. Instead of users transforming data into clean, multi-dimensional structures at the database level, complex SQL provides equivalent functionality (to a certain degree) at the query level. As a useful byproduct, this gives DWA-based analytic applications data-model independence, which is lost when analysts depend on remodeling data.

**The Data Warehouse Appliance Redefined**

Due to the trend among vendor offerings from whole-technology stacks to partial ones, a newly revised definition must encompass DWAs that comply with the original definition (from Netezza and DATAallegro), as well as the newer software appliances (from Greenplum, Kognito, ParAccel, Vertica, and so on). A truly comprehensive definition will also include variations on the DWA theme (such as the appliances mentioned from Dataupia, Cognos, and SAP). Furthermore, hardware/software bundles assembled for data warehousing (though using mostly general purpose components) share many characteristics and benefits with DWAs, so these should be mentioned anytime DWAs come up. (See Table 1, next page, for a summary of vendors and products.)

Vendor products aside, the user community continues to redefine how it uses data warehouse appliances. As discovery-driven analytics become more common and more mission critical, users are in more dire need for a data warehouse platform that can respond quickly (with little or no tuning) to ad hoc and/or complex queries against multi-terabyte data sets of less-than-ideal structure and quality. To get the analytic databases they need, users will probably continue the trend of the SOS—systems on the side that augment the analytic capabilities of an enterprise data warehouse environment.
Note that the next generation of data warehouse appliances is about **diversity**. Adjust your definition of DWAs to include both whole technology stack and partial technology stack approaches. Don’t forget that many hardware/software bundles have characteristics and benefits similar to DWAs, as do the BI accelerators that are packaged as appliances.

**Expect the definition of the data warehouse appliance to evolve—again.** DWAs are still new, and their possibilities are still being explored.

**Be open to alternative DBMSs for a DW platform, including open source and columnar databases.** Otherwise, you exclude most data warehouse appliances and some bundles. Likewise, be open to Linux, which is the most common operating system for DWAs and similar bundles.

**Know your requirements, and select a DW platform that matches them.** Don’t acquire a DWA based solely on its compelling low cost or perky query performance. Know the DWA sweet spots and watch for these while gathering data warehouse platform requirements.

**Don’t replace or overtax your EDW.** It’s still the single version of the truth for most reporting and analysis. Help it play that role by off-loading taxing analytic applications to DWAs and other SOS platforms.

**Your evaluation list should include DWAs, software appliances, and similar bundles.** After all, these are now established data warehouse platforms, along with more traditional platform components, such as relational databases and the usual hardware servers.

---

**Summary of Data Warehouse Appliance Vendors**

<table>
<thead>
<tr>
<th>Data Warehouse Appliances (DWAs)</th>
<th>Whole Technology Stack</th>
<th>Partial Technology Stack (or “Software Appliances”)</th>
<th>Miscellaneous</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Original DWAs:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Netezza</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DATAAllegro (now Microsoft)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Precedents:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Teradata</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sequent, White Cross</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>New:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HP Oracle Database Machine</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exadata</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Relational Databases:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Greenplum</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kognitio</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Columnar Databases:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ParAccel</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sybase IQ</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vertica</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>General Data Appliance:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dataupia Sartori Server</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Add-on Accelerators:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cognos Now!</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SAP BI Accelerator</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Teradata Accelerate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Analytic Databases:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Illuminate, Aster Data</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Related Products</th>
<th>Server/Database Bundles:</th>
<th>Open Source Databases:</th>
<th>Commodity Hardware:</th>
</tr>
</thead>
<tbody>
<tr>
<td>HP Neoview</td>
<td></td>
<td></td>
<td>Dell, HP</td>
</tr>
<tr>
<td>IBM InfoSphere on IBM</td>
<td></td>
<td></td>
<td>Cisco, EMC</td>
</tr>
<tr>
<td>Oracle on Sun or HP</td>
<td></td>
<td></td>
<td>Ingres</td>
</tr>
<tr>
<td>Sybase IQ on Sun or IBM</td>
<td></td>
<td></td>
<td>PostGres</td>
</tr>
<tr>
<td>Many others</td>
<td></td>
<td></td>
<td>MySQL</td>
</tr>
<tr>
<td><strong>Open Source Op Sys:</strong></td>
<td></td>
<td></td>
<td>LINUX (various vendors)</td>
</tr>
<tr>
<td><strong>Open Source Databases:</strong></td>
<td></td>
<td></td>
<td>Many others</td>
</tr>
<tr>
<td>Infobright</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ingres</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PostGreys</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MySQL</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1. The vendors and products listed here are representative, and the list is not meant to be comprehensive. Source: TDWI Research 2009.

**Recommendations**

* Philip Russom is the senior manager of TDWI Research at The Data Warehousing Institute (TDWI), where he oversees many of TDWI’s research-oriented publications, services, and events. He’s been an industry analyst researching BI issues at Forrester Research, Giga Information Group, and Hurwitz Group. You can reach him at prussom@tdwi.org.

Revolution Wireless Dials in Key Company Strategies, Competitive Advantages with Business Intelligence Solutions

Commentary by George Mehok
Chief Information Officer, Revolution Wireless

The Company
Founded in 2005 and headquartered in Independence, OH, Revolution Wireless is a regional provider of wireless communications. The company leverages its high-quality, all-digital wireless network platform to offer truly unlimited calling at industry-leading pricing. Enjoying tremendous growth, Revolution has more than 400 employees and retail locations in Ohio and Indiana.

Challenge
Revolution’s strategy for winning in a highly competitive climate relies on proactive churn management and on creating a corporate culture stressing performance management, accountability, and internal competitiveness. Revolution sees business intelligence as crucial to supporting these strategies.

As is typical for fast-growing enterprises, Revolution was plagued with disjointed, frequently inconsistent information “silos.” “But to measure performance adequately,” says George Mehok, CIO, “we needed to analyze consistent data along various dimensions, including store, region, associate, project, and others.”

To meet these needs, the Revolution IT team began to engineer a single, reliable source of truth: a data warehouse and BI solution to be accessible to as many business users as possible, company wide. Dashboards would facilitate performance monitoring and operational reporting to track key performance indicators (KPIs), while analytics would help uncover churn factors.

Choosing Business Intelligence Technology
Because several operational applications were Oracle-based, Oracle was chosen as the data warehouse platform. The team evaluated business intelligence solutions from four proprietary vendors and a pair of open source providers; requirements included solid integration/ETL tools, flexible reporting, easy-to-build dashboards, and ad hoc query features.

Self-service usability was also critical, since wide-deployment goals demanded that end users be able to define reports and dashboards without help from IT. Pricing was key, and because Revolution was already using other open source technologies such as Apache, Linux, and PHP, an open source BI solution was determined to be a good fit, promising flexibility as new technology initiatives arose later.

Solution
Revolution chose the commercial open source Jaspersoft Business Intelligence Suite for its ease of use; its low-cost, per-server subscription pricing model; and its modern, open architecture and support for Web 2.0 user interfaces. Rich APIs and easy access to source code would facilitate fast, easy customization, and user adoption would be enhanced by flexible output options including mobile device–friendly reporting.

Other positives included the consistent high marks given to Jaspersoft in independent industry surveys, as well as the community support, ease of integration, and ongoing source code access advantages of the open source software model.

Except for data extraction from a few Oracle-based operational data sources, all tasks—ETL, reporting, report design, report scheduling, dashboarding, and ad hoc query—are managed with Jaspersoft tools. To create a pervasive BI environment, Revolution has also initiated data governance processes and has incorporated Web 2.0 UI tools.

“We’re visibly moving the company strategy forward. We’ve built a strategic asset, and we’re a better company because of this work. And that makes everyone feel good.”

George Mehok
Chief Information Officer, Revolution Wireless

Results
More than 100 employees use Jaspersoft to access Revolution’s data warehouse, enabling the company to boost market share and improve sales effectiveness—and to enjoy other interesting advantages.

Historical sales information in the data warehouse, combined with demographic data from state government sources, has fueled traffic studies for optimizing store locations. Real-time monitoring, coupled with reports and dashboards, enables competitive sales analysis across stores, regions, and even sales associates. It also facilitates quick responses when unusual patterns suggest problems or opportunities.

“We’re also having good success reinforcing the cultural changes we want,” says Mehok. “Just about every report or dashboard with a KPI or other metric compares performance to targets—and that keeps us all focused on accountability.”

One final reward: The new data warehouse and BI capabilities are viewed as a strategic asset—a positive for the IT team along with the rest of the enterprise.

For a white paper on this topic, click here and choose the title “Open Source Business Intelligence Costs and Benefits: What You Need to Know.” For more information about Jaspersoft, click here.
Data warehouses have to manage more and more information each day, while business users continue to expect consistently fast query response times. Getting the right balance of server, storage, networking, and intelligent software to deliver on these expectations can be a challenge. Of equal importance is finding a solution that does not compromise the scalability, reliability, security, and support for mixed workloads, analytics, and data mining that are crucial to the success of any data warehouse.

**What to Consider**

Ignoring the basics of properly architecting a data warehouse can result in an expensive failure. Here are some basics to consider when architecting your data warehouse.

**System performance and scalability.** Start by defining your user population’s requirements in terms of complexity and frequency of queries, both today and in the future. Don’t forget to factor in how your data volume is expected to grow, the frequency of data loads, and the mix of workloads, or query performance and system scalability issues will arise. To avoid these issues, you need a data warehouse that can easily scale out data storage, network bandwidth, and processing capacity.

**Required analysis capabilities.** Next, establish your analytic requirements. Most businesses need more than regular reports and ad hoc query analysis. For example, data mining engines analyze data to deliver forward-looking insights using sophisticated predictive techniques. Likewise, OLAP engines can deliver advanced analytics such as forecasting and what-if analysis very quickly. Architectural complexity is greatly reduced if these engines are hosted within the data warehouse database itself, eliminating the need for multiple repositories and engines.

**User service-level agreements.** In addition to acceptable performance criteria, users are increasingly dependent on data warehouses to make timely decisions in a secure manner; downtime or loss of access to information is unacceptable. It’s important to design a system based on proven high-availability technologies that can protect your data warehouse from planned and unplanned downtime. Plus, you need an infrastructure that fully protects sensitive information throughout the enterprise. Using high availability and data protection features in your data warehouse, rather than manually integrating third-party tools, can dramatically reduce implementation costs and the cost of meeting users’ service-level objectives.

**Time to market.** Finally, you need to map the skills and resources required to implement and manage your data warehouse today and in the future to what you already have. Where in-house resources may not be available, it’s important to build your data warehouse using industry-standard components that can be supported by resources and skills that are readily available. This helps reduce your time-to-market and ongoing management costs and protects your long-term data warehousing investment.

**Next Steps**

Once you’ve identified a list of solution providers that meet your requirements, you should talk to reference customers. Also, consider running benchmark tests, or proof of concepts, but be sure to set aside reasonable timelines in your project plans for these activities. And always use real source data from operational systems; using data subsets will not reflect your go-live deployment, and will put project success at risk.

**What Oracle Offers**

Oracle leads the database and data warehouse markets with innovative products, including Oracle Database 11g and the Oracle Database Machine. In addition to industry-leading performance and scalability, Oracle Database 11g offers integrated analytics on reliable, low-cost grids. The Oracle Database Machine is a complete data warehousing grid—built using Oracle Database 11g and industry-standard servers, storage, and networking components—that delivers extreme performance for large-scale data warehouses. It’s a fast, reliable, and secure system that can easily scale to meet the complex reporting and analytics needs of the most demanding organizations.

For more information about Oracle Corporation, click here.
KPI Components

**Metrics versus KPIs.** The term *metric* refers to measurements of business activity. But in a performance management system, we want to do more than just measure business activity; we want to measure performance aligned with business strategy.

To distinguish between run-of-the-mill metrics and strategically aligned metrics, cognoscenti use the term *key performance indicator*, or KPI. The only difference between a metric and a KPI is that a KPI embodies a strategic objective and measures performance against a goal. The goals attached to a KPI are multidimensional: they have ranges that are encoded in software, a time frame by which the goals must be achieved, and a benchmark against which the goals are compared. (See Table 1.)

**Goals and Targets.** The goals associated with KPIs are known as *targets* because they specify a measurable outcome rather than a conceptual destination. Ideally, executives, managers, and workers collectively set targets during strategic, planning, or budget discussions. Such collaboration ensures buy-in and more accurate targets. Targets can also be set by a KPI team charged with translating strategic objectives into a performance plan.

Since KPIs derive much of their impact from the targets associated with them, it’s important to differentiate between types of targets. Ranga Bodla, senior director of enterprise performance software at SAP, defines five types of targets:

1. **Achievement:** Performance should reach or exceed the target. Anything over the target is valuable but not required. Examples include revenues and satisfaction.

2. **Reduction:** Performance should reach or be lower than the target. Anything less than the target is valuable but not required. Examples include overtime and attrition.

3. **Absolute:** Performance should equal the target. Anything above or below is not good. Examples include in-stock percentage and on-time delivery.

4. **Min/max:** Performance should be within a range of values. Anything above or below the range is not good. Example: mean time between repairs.

5. **Zero:** Performance should equal zero, which is the minimum value possible. Examples include employee injuries and product defects.

**Ranges.** To help workers gauge their performance, most organizations stratify targets; that is, they divide the target into ranges, usually percentage increments above and below the target. (Think of the concentric circles around a bull’s-eye in an archery target.) Organizations can create any number of ranges around a target, although most use three: above target, on target, or below target.

**Encodings and Thresholds.** The next step is to encode ranges into graphical displays on a performance dashboard or report. It’s important that organizations apply a consistent encoding...
or scoring system to their KPIs no matter what type of targets they use. That way, when users look at a KPI display within a dashboard, all the conditional formatting (i.e., color coding) will connote the same performance values, according to Bodla. For instance, above target, on target, or below target ranges easily translate into red, yellow, and green color encodings. The boundary line between ranges is called a threshold. When performance exceeds a threshold, the graphical representation of the KPI changes, say from yellow to green. Threshold-based encodings make it easy for users to assess the status of key business processes or projects at a glance, which is the major reason performance dashboards are so popular.

Although the use of three ranges is common, I once saw an operations group use 16 ranges and encoding schemes. Some organizations use multiple ranges to emphasize “stretch” or visionary goals that motivate employees to achieve above the norm and think outside the box. Others want to track granular gradations in performance, especially when they’ve attached monetary incentives to achievement levels.

Not all organizations encode target ranges using color; some use a variety of graphical devices, such as bullet graphs or other icons. For example, a beer manufacturer might use graphics of beer mugs that are empty, half-full, or full to portray performance against goals.¹ The number of target ranges and choice of graphics depends on a group’s culture and preferences.

Encoding Nuances. Encoding performance is not always straightforward. Many organizations display only a handful of KPIs in an executive scorecard. As a result, those KPIs may be averages or aggregates based on multiple lower-level KPIs. The problem with averages and aggregates is that they may show above-target performance even when one or more of the lower-level metrics registers abysmal performance scores.

To avoid such dilution, architects apply Boolean rules to the thresholds and encodings. For instance, they might write a rule that says if one business unit scores in the bottom range, then mark the overall score as below average even if the rest of the associated KPIs are above average. Or architects may assign double or triple weights to the scores in the bottom range so they have a greater impact on the higher-level KPIs.

Time Frames. Performance targets have time frames, which affects how KPIs are calculated and displayed. Many organizations establish annual targets for key processes and initiatives. To keep employees on track to achieve those long-term targets, many organizations divide time frames into intervals that are measured on a more frequent basis. For instance, a group may divide the annual target to improve
customer satisfaction from 60% to 64% into four quarterly intervals with 1% targeted improvement each quarter. In some cases, such as a retail environment affected by seasonal shopping, groups may backweight the targets toward the end of the year, since most sales occur during the December holiday season. (See Figure 1, previous page.)

**Benchmarks.** Finally, KPI targets are measured against a benchmark that becomes the starting point for improving performance. Typically, the benchmark is last year’s output. So, for example, a sales team may need to increase sales by 15% compared to last year. Or the benchmark could be an external standard, such as the performance level of an industry leader. Here, a company might want to set a goal of closing the gap in market share with its closest rival by 50% next year.

In some cases, the benchmark may be completely arbitrary. This often occurs with visionary goals, where an executive mandates a seemingly impossible goal to achieve. Like John F. Kennedy’s famous imperative to send a man to the moon by the end of a decade, visionary goals force employees to think outside existing norms and processes to achieve the desired outcome.

**Reading KPIs**

Reading KPIs should be straightforward. Users should be able to look at a visual display that has been properly encoded and know instantly whether a process or project is on track. But we find that many employees don’t know how to read KPI displays—or more alarming—how to interpret KPIs. (The latter involves a degree of analytical literacy that we’ll address later.)

**Seven Attributes.** To assist users with interpreting KPIs, a good performance dashboard displays seven attributes for each (see Figures 2a and 2b, next page):

1. **Status**
2. **Trend**
3. **KPI name**
4. **Actual value**
5. **Target value**
6. **Variance**
7. **Variance percentage**

**Status** measures performance against the target and is usually shown with a stoplight. **Trend** measures performance against the prior interval or another time period and is often displayed using arrows or trend lines. The KPI name is usually hyperlinked so users can click on it to view a time series chart. The actual and target values are self-explanatory and usually displayed with text. **Variance** measures the gap between actual and target and is displayed using text or a micro bar chart or bullet chart. Variance percentage divides the variance against the target. These seven attributes combine to provide valuable insight into the state of performance.

Status is the attribute most people respond to first, since it’s usually displayed as a stoplight or gauge. A red stoplight on a dashboard jumps out at the viewer, who feels impelled to take notice. From there, users will examine variance and variance percentage to help them understand with precision just how bad (or good) the situation is. They then look at the trend, which signifies whether the situation is new or old, improving or declining. In essence, status signals a problem, variance describes its extent, and trend defines its movement.

But collectively, these attributes can give mixed signals. For example, if a KPI’s status is red but its trend is positive, what should a user think? Perhaps someone already spotted the problem and applied a remedy that is now working. Conversely, what if a KPI’s status is green but its trend is downward? A green light indicates that performance is above average, but shouldn’t the user take action now before the green light turns yellow?

Interpreting KPI attributes gets even more complex when the targets change each period or interval. For instance, if a KPI’s status is green and its trend is positive, but its variance is increasing, should the user be alarmed? This combination of signals indicates that performance is improving, but not fast enough to meet accelerated interim targets. In this case, many companies plot percentage variance, not actual values, to highlight whether the gap between actual and target is growing or shrinking over time. This takes some of the guesswork out of interpreting KPIs, which is especially helpful for new employees or managers who’ve been assigned to a new area.

**Types of KPIs**

**Outcome versus Driver KPIs.** There are two fundamental types of KPIs: outcomes and drivers. Outcome KPIs—sometimes known as lagging indicators—measure the output of past activity. They are often financial in nature, but not always. Examples include revenues, margins, return on equity, customer satisfaction, and employee retention.

On the other hand, driver KPIs—sometimes known as leading indicators or value drivers—measure activities that
have a significant impact on outcome KPIs. These KPIs measure activity in its current state (number of sales meetings today) or a future state (number of sales meetings scheduled for the next two weeks). The latter is more powerful, since it gives individuals and their managers more time to adjust behavior to influence a desired outcome.

There is confusion about the difference between these two types of KPIs. (Some industry experts add performance indicators, key risk indicators [KRIs], diagnostic indicators, and other types of KPIs, which clouds the issue.) Truth be told, there often isn’t a lot of difference. “One man’s outcome measure is another man’s value driver,” says Neal Williams, founder of Corda and currently head of its professional services. “An outcome KPI in one dashboard could be a driver KPI in another.”

Creating KPIs. Most organizations find it easy to create outcome KPIs because they measure the desired state of a strategic objective. In fact, most companies already track outcome KPIs in management reports or dashboards, although most are slanted to financial objectives. Not surprisingly, the majority of KPIs (56%) are outcome or lagging indicators, although a majority (57%) are non-financial measures, according to our research. (See Figure 3, next page.)

On the other hand, many organizations struggle to define accurate drivers of future performance. It often takes months or years of trial and error to find metrics that correlate with future outcomes. A place to start is a brainstorming session in which a group identifies desired outcomes and then asks, “What behaviors or activities will drive or lead to that outcome?” Each time the group identifies a new driver, it should ask the same question: “What behaviors or activities will produce that result?” After several iterations, the group will find one or two root causes that have the most significant impact on the desired outcome.

Organizations can also conduct a statistical analysis to correlate outcomes with driver KPIs. For example, one online service provider identified two KPIs that correlate with the ability of telemarketers to meet daily sales quotas: the amount of time they spend on the phone with customers and the number of clients they speak with each day. The company...
operational metrics should drive higher-level KPIs, or it isn’t worth the time and energy to track them.

**Qualitative and Quantitative KPIs.** Another distinction between KPIs is that some are based on quantitative data, while others are based on qualitative or subjective data. Quantitative data measures activity by counting, adding, or averaging numbers. Operational systems that manage inventory, supply chain, purchasing, orders, accounting, and so on all gather quantitative data used by KPIs. Financial KPIs are based on quantitative data, as are employee injuries, number of training classes, and so on. Quantitative data forms the backbone of most KPIs.

But qualitative KPIs are just as important. The most common ones gauge customer or employee satisfaction through surveys. While the survey data itself is quantitative, the measures are based on a subjective interpretation of a customer’s or employee’s opinion on various issues. These opinions can help explain why performance is dropping when all other indicators seem fine. Many companies use customer satisfaction KPIs to refine products and optimize processes.

“Our annual customer satisfaction survey measures customer perception of Cisco’s performance on all aspects of doing business with us,” says David Hsiao, director of quality data infrastructure at Cisco Systems. “We then analyze this perception data along with internal process measures, field experience, and financial data to identify linkages. Regression analysis then identifies the key drivers that impact customer loyalty, revenue, and profitability per geographic theatre, vertical market, customer account, and product. This provides invaluable insights to us on how to run our business.”

**Wayne W. Eckerson** is the director of TDWI Research at The Data Warehousing Institute. He 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.

This article was excerpted from the full, 32-page report, Performance Management Strategies: How to Create and Deploy Effective Metrics. 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; Corda Technologies; IBM Corporation; MicroStrategy; and Pentaho Corporation.
INTRODUCTION TO OPERATIONAL DATA INTEGRATION

By Philip Russom

The amount and diversity of work done by data integration specialists has exploded since the turn of the twenty-first century. Analytic data integration continues to be a vibrant and growing practice that’s applied most often to data warehousing and business intelligence initiatives. But a lot of the growth comes from the emerging practice of operational data integration, which is usually applied to the migration, consolidation, or synchronization of operational databases, plus business-to-business data exchange. Analytic and operational data integration are both growing; yet, the latter is growing faster in some sectors.

But growth comes at a cost. Many corporations have staffed operational data integration by borrowing data integration specialists from data warehouse teams, which puts important BI work in peril. Others have gone to the other extreme by building new teams and infrastructure that are redundant with analytic efforts. In many firms, operational data integration’s contributions to the business are limited by legacy, hand-coded solutions that are in dire need of upgrade or replacement. And the best practices of operational data integration on an enterprise scale are still coalescing, so confusion abounds.

The purpose of this report is to identify the best practices and common pitfalls involved in starting and sustaining a program for operational data integration. The report defines operational data integration in terms of its relationship to other data integration practices, as well as by its most common project types. Along the way, we’ll look at staffing and other organizational issues, followed by a list of technical requirements and vendor products that apply to operational data integration projects.

The Three Broad Practice Areas of Data Integration

TDWI’s position is that diverse data integration practices are distinguished by the larger technical projects or business initiatives they support. So, this report defines data integration practices by their associated projects and initiatives. Figure 1, next page, summarizes these projects, initiatives, and practices (plus relationships among them) in a visual taxonomy. There are three broad practice areas for data integration:

• **Analytic data integration (AnDI)** is applied most often to data warehousing (DW) and business intelligence (BI), where the primary goal is to extract and transform data from operational systems and load it into a data warehouse. It also includes related activities like report and dashboard refreshes and the generation of data marts or cubes. Most AnDI work is executed by a team set up explicitly for DW or BI work.

• **Operational data integration (OpDI)** is more diverse and less focused than AnDI, making it harder to define. For this reason, many of the users TDWI interviewed for this report refer to OpDI as the “non-data-warehouse data integration work” they do. To define it more positively, OpDI is: “the exchange of data among operational applications, whether in one enterprise or across multiple ones.” OpDI involves a long list of project types, but it usually manifests itself as projects for the migration, consolidation, collocation, and upgrade of operational databases. These projects are usually considered intermittent work, unlike the continuous, daily work of AnDI. Even so, some OpDI work can also be continuous, as seen in operational database synchronization (which may operate 24x7) and business-to-business data exchange (which is critical to daily operations in industries as diverse as manufacturing and retail or financials and insurance). OpDI work is regularly assigned to the database administrators and application developers who work on the larger initiatives with which OpDI is associated. More and more, however, DW/BI team members are assigned OpDI work.

• **Hybrid data integration (HyDI)** practices fall in the middle ground somewhere between AnDI and OpDI practices. HyDI includes master data management (MDM) and similar practices like customer data integration (CDI) and product information management (PIM). In a lot of ways, these are a bridge between analytic and operational practices. In fact, in the way that many organizations implement MDM, CDI, and PIM, they are both analytic and operational.

---

1 For another definition of operational data integration, see the article “Operational Data Integration” by Philip Russom in TDWI’s What Works, Volume 23. www.tdwi.org/publications/whatworks
As a quick aside, let’s remember that data integration is accomplished via a variety of techniques, including enterprise application integration (EAI), extract, transform, and load (ETL), data federation, replication, and synchronization. IT professionals implement these techniques with technique-specific vendor tools, hand coding, functions built into database management systems (DBMSs) and other platforms, or all of these. And all the techniques and tool types under the broad rubric of data integration operate similarly in that they copy data from a source, merge data coming from multiple sources, and alter the resulting data model to fit the target system that data will be loaded into. Because of the similar operations, industrious users can apply just about any technique (or combination of these) to any data integration implementation, initiative, project, or practice—including those for OpDI.

The Three Main Practice Areas within OpDI

Now that we’ve defined AnDI, HyDI, and OpDI, we can dive into the real topic of this report: OpDI and its three main practice areas, as seen in the bottom layer of Figure 1.

DATA MIGRATION

Although it’s convenient to call this practice area “data migration,” it actually includes four distinct but related project types. Data migrations and consolidations are the most noticeable project types, although these are sometimes joined by similar projects for data collocation or database upgrade. Note that all four of these project types are often associated with applications work. In other words, when applications are migrated, consolidated, upgraded, or collocated, the applications’ databases must also be migrated, consolidated, upgraded, or collocated.

Migration. Data migrations typically abandon an old platform in favor of a new one, as when migrating data from a legacy hierarchical database platform to a modern relational one. Sometimes the abandoned database platform isn’t really a “legacy”; it simply isn’t the corporate standard.

Consolidation. Many organizations have multiple customer databases that require consolidation to provide a single view of customers. Data mart consolidation is a common example in the BI world. And consolidating multiple instances of a packaged application into one involves consolidating the databases of the instances.

Upgrade. Upgrading a packaged application for ERP or CRM can be complex when users have customized the application and its database. Likewise, upgrading to a recent version of a database management system is complicated when users are two or more versions behind.

Collocation. This is often a first step that precedes other data migration or consolidation project types. For example, you might collocate several data marts in the enterprise data warehouse before eventually consolidating them into the warehouse data model. In a merger and acquisition, data from the acquired company may be collocated with that of the acquiring company before data from the two are consolidated.

These four data migration project types are related because they all involve moving operational data from database to database or application to application. They are also related because users commonly apply one or more of these project types together. Also, more and more users apply the tools and techniques of data integration to all four. But beware, because migration projects are intrusive—even fatal—in that they kill off older systems after their data has been moved to another database platform.
DATA SYNCHRONIZATION
Killing off a database or other platform—the way data migrations and consolidations do—isn’t always desirable or possible. Sometimes it’s best to avoid the risk, cost, and disruption of data migration and leave redundant applications and databases in place. When these IT systems share data in common—typically about business entities like customers, products, or financials—it may be necessary to synchronize data across the redundant systems so that the view of these business entities is the same from each application and its database. For example, data synchronization regularly syncs customer data across multiple CRM and CDI solutions, and it syncs a wide range of operational data across ERP applications and instances. Furthermore, when a data migration project moves data, applications, and users in multiple phases, data sync is required to keep the data of old and newly migrated systems synchronized.

Note that true synchronization moves data in two or more directions, unlike the one-way data movement seen in migrations and consolidations. When each database involved in synchronization is subject to frequent inserts and updates, it’s inevitable that some data values will conflict when multiple systems are compared. For this reason, synchronization technology must include rules for resolving conflicting data values.

Hence, synchronization is a distinct practice that’s separate from migration, consolidation, and other similar OpDI practices, because—unlike them—it leaves original systems in place, is multi-directional, and can resolve conflicting values on the fly. Furthermore, migrations and consolidations tend to be intermittent work, whereas data synchronization is a permanent piece of infrastructure that runs daily for years before reaching the end of its useful life.

BUSINESS-TO-BUSINESS (B2B) DATA EXCHANGE
For decades now, partnering businesses have exchanged data with each other, whether the partners are independent firms or business units of the same enterprise. A minority of corporations use applications based on electronic data interchange (EDI) operating over value-added networks (VANs). These are currently falling from favor, because EDI is expensive and limited in the amount and format of data it supports. In most B2B situations, the partnering businesses don’t share a common LAN or WAN, so they share data in an extremely loosely coupled way, as flat files transported through file transfer protocol (FTP) sites. Flat files over FTP is very affordable (especially compared to EDI), but rather low-end in terms of technical functionality. In recent years, some organizations have upgraded their B2B data exchange solutions to support XML files and HTTP; this is a small step forward, leaving plenty of room for improvement.

B2B data exchange is a mission-critical application in industries that depend on an active supply chain that shares a lot of product information, such as manufacturing and retail. It’s also critical to industries that share information about people and money, like financials and insurance. Despite being mission-critical, B2B data exchange in most organizations remains a low-tech affair based on generating and processing lots of flat files. According to users TDWI Research interviewed for this report, most B2B data exchange solutions (excepting those based on EDI) are hand-coded legacies that need replacing. Most of these have little or no functionality for data quality or master data management, and they lack any recognizable architecture or modern features like Web services. Hence, in the user interviews, TDWI found that organizations are currently planning their next-generation B2B data exchange solutions, which will be built atop vendor tools for data integration with support for data quality, master data, services, business intelligence, and many other modern features.

USER STORY
DIFFERENT INDUSTRIES HAVE DIFFERENT OPDI NEEDS.

“I worked in financial services for years, where I was unpredictably bombarded with system migration and consolidation work, as the fallout of periodic mergers and acquisitions. Now that I work in e-commerce, my operational data integration work focuses on upgrades and restructuring of our e-commerce applications, largely to give them greater speed and scalability.”
Why Care about Operational Data Integration Now?

There are many reasons organizations need to revisit their OpDI solutions now to be sure they address current and changing business requirements:

- **OpDI is a growing practice.** More organizations are doing more OpDI; it is an increasing percent of the workload of data integration specialists and other IT personnel. Despite the increase in OpDI work, few organizations are staffing it appropriately.

- **OpDI solutions are in serious need of improvement or replacement.** Many are hand-coded legacies that need to be replaced by modern solutions built atop vendor tools.

- **OpDI solutions tend to be feature poor.** They need to be augmented with functions they currently lack for data quality, master data, scalability, maintainable architecture, Web services, and modern tools for better developer productivity and collaboration.

- **OpDI and AnDI have different goals and sponsors.** Hence, the two have different technical requirements and organizational support. Don’t assume you can do both with the same team, tools, and budget.

- **OpDI addresses real-world problems and supports mission-critical applications.** So you should make sure it succeeds and contributes to the success of the initiatives and projects it supports.

In short, operational data integration is a quickly expanding practice that user organizations need to focus on now—to foster its growth, staff it properly, provide it with appropriate technical infrastructure, and assure collaboration with other business and technology teams. The challenge is to develop the new frontier of operational data integration without scavenging unduly from similar efforts in analytic data integration.

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, 32-page report, *Operational Data Integration: A New Frontier for Data Management*. You can download this and other TDWI Research free of charge at www.tdwi.org/research/reportseries.

The report was sponsored by DataFlux, expressor software, Golden-Gate Software, IBM, Informatica Corporation, SAP BusinessObjects, Silver Creek Systems, Sybase, and Talend.
The following solution providers have shared their data integration and data warehousing stories and successes, technology insights, and the lessons they have learned for *What Works in Data Integration*.

**Baseline Consulting**

15300 Ventura Boulevard, Suite 523
Sherman Oaks, CA 91403
818.906.7638
Fax: 818.907.6012
inquiries@baseline-consulting.com
www.baseline-consulting.com

Baseline Consulting is a management and technology consulting firm that provides data integration and business analytic services to help companies enhance the value of enterprise data and improve the performance of their business. Baseline’s proven, structured approaches uniquely position us to help clients achieve self-sufficiency in designing, delivering, and managing data as a corporate asset.

**Tidal Software (now a part of Cisco)**

2100 Geng Road, Suite 210
Palo Alto, CA 94303
877.558.4325
Fax: 650.475.4700
info@tidalsoftware.com
www.tidalsoftware.com

Cisco Systems, Inc. is the worldwide leader in networking for the Internet. Today, networks are an essential part of business, education, government, and home communications, and Cisco Internet Protocol–based (IP) networking solutions are the foundation of these networks.

Cisco hardware, software, and service offerings are used to create Internet solutions that allow individuals, companies, and countries to increase productivity, improve customer satisfaction, and strengthen competitive advantage. The Cisco name has become synonymous with the Internet, as well as with the productivity improvements that Internet business solutions provide. At Cisco, our vision is to change the way people work, live, play, and learn.
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).

GoldenGate Software Inc. is a leading provider of real-time data integration and high availability solutions for improving the availability, accessibility, and performance of critical data across heterogeneous enterprise IT environments. GoldenGate’s real-time data integration capabilities offer low-impact, flexible transactional change data capture that reduces data latency and enables 24/7 access to operational data. GoldenGate can provide continuous data feeds directly to various target systems, or integrate with ETL and messaging environments. More than 500 customers use GoldenGate in over 4,000 solution implementations worldwide.

Harte-Hanks Trillium Software
25 Linnell Circle
Billerica, MA 01821
978.436.8900
Fax: 978.670.5793
trilliumsoftware@trilliumsoftware.com
www.trilliumsoftware.com

Enterprise data quality ensures that accurate, consistent, and complete information is always available to drive higher productivity, decrease costs, increase revenues, and improve ROI. Trillium Software lets businesses implement a disciplined approach to total data quality across the enterprise and eradicate data quality defects.

Harte-Hanks Trillium Software, the global leader in enterprise data quality, provides technologies and services that deliver global data profiling, cleansing, enhancement, linking, and governance for CRM, ERP, supply chain management, data warehouse, e-business, and other enterprise applications.

The Trillium Software System®:

• Allows collaboration across business and IT resources

• Enables discovery and profiling of large data volumes

• Cleanses, standardizes, matches, and supplements any data type including customer, product, and financial data

• Monitors business rules and data quality metrics through a customizable data quality dashboard
Initiate Systems, Inc.

200 West Madison, Suite 2300
Chicago, IL 60606
800.992.0300
Fax: 312.759.5026
info@initiate.com
www.initiate.com

Initiate Systems enables organizations to share critical data assets. Multinational corporations, healthcare organizations, and government agencies rely on Initiate® software and services to deliver complete, accurate, and real-time views of data spread across multiple systems, domains, or databases. Initiate’s multidomain master data management (MDM) solutions facilitate interoperability, information sharing, entity resolution, and data quality to increase revenues, reduce costs, and mitigate risks.

Praised by analysts for having the most productive live customers and the most significant research and development gains, Initiate® transforms information into strategic advantage. For more information, visit www.initiate.com.

iWay Software

Two Penn Plaza
New York, NY 10121
212.330.1700
Fax: 212.564.1726
info@iwaysoftware.com
www.iwaysoftware.com

iWay Software’s integration methodology is simple: It uses a single, integrated set of graphical design tools to assemble powerful prebuilt components for enterprise-class business-to-business (B2B), business process automation (BPA), or enterprise information management (EIM) integration scenarios—without the use of custom code. Integration configurations can be deployed in a stand-alone manner to any environment supporting a JVM—or to Web application servers from any vendor. iWay delivers the fastest, most cost-effective, and simplest way to integrate and streamline critical business processes. iWay Software’s vendor-, platform-, and protocol-neutral solutions achieve the promise of SOA: true reusability.

Jaspersoft

539 Bryant Street, Suite 100
San Francisco, CA 94107
415.348.2300
Toll Free: 888.399.2199
Fax: 415.281.1987
www.jaspersoft.com

Jaspersoft’s open source business intelligence suite is the world’s most widely used BI software, with more than 8 million total downloads worldwide and more than 10,000 commercial customers in 96 countries. The company’s Jaspersoft Business Intelligence Suite provides a Web-based, open, and modular approach that addresses the next generation business intelligence needs of the enterprise. Jaspersoft’s software is rapidly updated by a community of more than 90,000 registered members working on more than 350 projects, which represents the world’s largest business intelligence community. More information is available at www.jaspersoft.com and www.jasperforge.org.
Since 1985, Melissa Data has been providing scalable data quality and enrichment solutions that help SMB to Fortune 500 organizations verify, standardize, and optimize their data assets across the enterprise. Our multiplatform Data Quality Suite, Web services, and new Total Data Quality Integration Toolkit are employed by more than 5,000 customers to keep data clean for immediate ROI and significant cost savings. Melissa Data offers free trials, unlimited tech support, and a 30-day money back guarantee.

MicroStrategy, a global leader in business intelligence and performance management technology, provides reporting, analysis, and monitoring software that enables leading organizations to make better business decisions every day. Designed to support the most demanding business intelligence applications, MicroStrategy is ideal for enterprisewide BI standardization. Companies choose MicroStrategy for its advanced technical capabilities, sophisticated analytics, and superior data and user scalability. MicroStrategy is built from a single architectural foundation, making it the most integrated and efficient BI architecture available. With an intuitive Web interface, MicroStrategy enables business users to seamlessly access enterprise data for enhanced decision making. More information about MicroStrategy is available at www.microstrategy.com.

Oracle Corporation is the world’s largest enterprise software company, providing database, middleware, and collaboration products; enterprise business applications; application development tools; and professional services that help customers collaborate, grow, measure outcomes, and report results with confidence.

Oracle leads the database and data warehouse markets with innovative products including Oracle Exadata and Oracle Database 11g, which together provide a comprehensive platform for addressing today’s data warehousing and BI requirements. Oracle Exadata overcomes the limitations of conventional storage to improve query performance by 10x or more. Oracle Database 11g offers integrated analytics and embedded data quality on a reliable, low-cost grid infrastructure.

For more information, visit www.oracle.com.
Siperian, Inc.
100 Foster City Boulevard, 2nd Floor
Foster City, CA 94404
1.866.SIPERIAN
info@siperian.com
www.siperian.com

Siperian's master data management (MDM) software platform provides a reliable and unified view of critical customer, product, supplier, and employee data to business users in Global 500 companies across 15 different industries. Through improved access to real-time information from distributed data sources, organizations can achieve higher profitability while reducing operational costs and improving compliance. Siperian's MDM Hub™ is a flexible and integrated software platform that significantly lowers total cost of ownership and delivers faster time to value and superior return on investment. Siperian meets the needs of data intensive companies by enabling them to improve customer relationship management, sales and marketing, regulatory compliance, and order-to-cash processes.

Talend
105 Fremont Avenue, Suite F
Los Altos, CA 94022
714.786.8140
Fax: 714.786.8139
info@talend.com
www.talend.com

Talend is the recognized market leader in open source data integration. 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, and load) for business intelligence and data warehousing, for migration, and for data quality management.

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.

Vision Solutions, Inc.
15300 Barranca Parkway
Irvine, CA 92618
949.253.6500
Toll Free: 800.683.4667
Fax: 949.253.6501
info@visionsolutions.com
www.visionsolutions.com

Vision Solutions, Inc. is the world’s leading provider of high availability, disaster recovery, and data management solutions for IBM® Power Systems. With a portfolio of innovative and trusted HA technologies from iTERA™, MIMIX®, and ORION Solutions™, Vision keeps critical business information continuously protected and available. Affordable and easy to use, Vision products ensure business continuity, increase productivity, reduce operating costs, and satisfy compliance requirements. Vision Solutions oversees a global network of partners and professionals to help customers achieve business goals. Privately held by Thoma Bravo, Inc., Vision Solutions is headquartered in Irvine, CA with offices worldwide. For more information, visit www.visionsolutions.com or call 800.957.4511.
BI Training Solutions:
As Close as Your Conference Room

We know you can’t always send people to training, especially in today’s economy. So TDWI Onsite Education brings the training to you. The same great instructors, the same great BI/DW education as a TDWI event—brought to your own conference room at an affordable rate.

It’s just that easy. Your location, our instructors, your team.

Contact Yvonne Baho at 978.582.7105 or ybaho@tdwi.org for more information. www.tdwi.org/onsite
About TDWI

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

MEMBERSHIP
www.tdwi.org/membership

In a challenging and ever-changing business intelligence and data warehousing environment, TDWI Membership offers a cost-effective solution for maintaining your competitive edge. TDWI will provide you with a comprehensive and constantly growing selection of industry research, news and information, online resources, and peer networking opportunities developed exclusively for its Members. TDWI offers a cost-effective way to keep your entire team current on the latest trends and technologies. TDWI’s Team Membership program provides significant discounts to organizations that register individuals as TDWI Team Members.

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.

BI SYMPOSIUMS
www.tdwi.org/ontheroad

TDWI BI Symposia offer the best of our Seminar Series combined with an exhibit area to keep you abreast of the latest in technology. As always, our vendor-neutral approach ensures that classes are loaded with objective, actionable information.

ONSITE EDUCATION
www.tdwi.org/onsite

TDWI Onsite Education 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.tdwi.org/cbip

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

WEBINAR SERIES
www.tdwi.org/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.