

Delivering information you can trust

March 2007



IBM **Information Management** software

Information integration: Challenges and benefits for enterprise application upgrades, consolidations and migrations

*Andrew Manby
Program Director
IBM Information Platform and Solutions*

Contents

- 2 Introduction**
- 4 Data integration poses specific challenges**
 - 5 *Developing a Center of Excellence for Data Integration*
 - 6 *Implementing master reference data architecture*
- 8 Current state versus desired state**
 - 9 *Reduce bottom-line costs, increase top-line growth*
- 10 Data integration helps DHL gain market share**
- 12 Moving toward an integrated platform**
 - 12 *Getting started*
- 14 Using an enterprise data integration platform**
- 18 Implementation methodology to create a scalable, repeatable solution**
- 20 EDF Energy dramatically reduces costs using data integration methodology**
- 21 Seven strategies to help reduce risk and costs**

Introduction

Many companies today are at an information technology (IT) crossroads. IT departments are being asked to deliver more with fewer resources. The success of many key business initiatives depends on significant IT intervention—initiatives such as the single view of a customer, strategic sourcing, enhanced inventory management and regulatory compliance for the Sarbanes-Oxley Act and other governmental reporting entities. The one common element for these initiatives is the requirement for consolidation flexibility around data.

This data consolidation is a significant challenge because of the fact that most organizations have many enterprise application instances—applications such as enterprise resource planning (ERP), customer relationship management (CRM), supply chain management (SCM), product lifecycle management (PLM) and supplier relationship management (SRM) from vendors such as SAP, Oracle (including PeopleSoft, JD Edwards and Siebel) and others. Additionally, many legacy and homegrown applications are scattered across the IT landscape. These applications were deployed over time—by country, by business unit, by department—and this situation was further complicated by corporate mergers and acquisitions that resulted in a combined entity that was running two or more of every application.

Harnessing this data and making sense of it often hinders the success of key business initiatives. In fact, the lack of accurate and consistent master reference data has become a huge problem and a top-of-mind business topic for line-of-business (LOB) and IT organizations alike. This lack of quality data is causing companies to explore new strategies for application upgrades, consolidation and migration—strategies that can have a positive impact on the business and enable top-line growth strategies.

In addition to the potential growth benefits from these initiatives, this complex application landscape situation also affects the bottom line. These applications are expensive—costs are multiplied by the number of application instances, the volume of software licenses, software maintenance renewal costs, hardware maintenance and upgrade costs, and staffing costs.

Further compounding these challenges is the recent consolidation of enterprise application vendors. Clearly, JD Edwards and PeopleSoft customers are faced with reviewing and evaluating their current enterprise application environment. Some questions that arise may include:

- *Do we stay with our current vendor and continue to upgrade existing applications or do we consider migrating to another vendor solution?*
- *If we migrate to another vendor, how quickly can we re-implement the ERP system? How will we transform our many business rules and all of the data?*
- *Or maybe we stay with what we have but consolidate other systems to address the disparate data issues that keep us from having harmonized master reference data to deliver a single view of a customer and power other key business initiatives.*

These issues are not unique to a certain type of business. Both large enterprises and small and medium businesses (SMBs) are faced with the challenge of making strategic decisions involving data.

Data integration poses specific challenges

Regardless of the current application environment, the issues surrounding data access and data quality still exist whether consolidating or upgrading applications, migrating to another vendor's system or simply keeping the status quo. The critical success factor lies in providing accurate and reliable data that can deliver a 360-degree view of a customer or a supplier, or prepare for Sarbanes-Oxley and other regulatory reporting burdens.

Some key data integration issues include:

- *Enterprise application source metadata is not easily assembled in one place to understand what is actually available. The mix may also include legacy sources, which often do not make metadata available through a standard application programming interface (API), if at all.*
- *Master reference data—names and addresses of suppliers and customers, part numbers and descriptions—differs across applications and/or duplicate sources of this data, which may be maintained between instances of a single application and between different ERP, CRM and SRM applications.*
- *Hundreds of extract/transform/load (ETL) jobs need to be written to move data from all the sources to the new target application.*
- *Data transformations are required before loading the data so it will fit into the new environment structures.*
- *New staff must be able to easily re-run and maintain the work as IT brings different teams online or as the rollout occurs in phases.*
- *The ability to handle large amounts of data that can be run through the process—and finish on time—is essential. Companies need the infrastructure to support running any of the transformation and data-matching routines on demand as well as in batch mode.*

Developing a Center of Excellence for Data Integration

Based on IBM discussions with customers and systems integrators who are frequently contracted to do this type of work, many companies do not know how to get started. They may be aware of some tools that can help or they may have used some in the past, but these challenges cannot be addressed solely based on a previously used point integration solution.

The most prevalent theory is that companies think they just need a single tool to accomplish these tasks—whether it is an enterprise application integration (EAI) tool or an ETL tool. Others believe they just need to write a few custom programs in a simple, well-known language.

In reality, a data integration project is much more than just extracting, transforming and loading data. For example, one customer who was evaluating ETL tools finally disclosed that the justification for the evaluation of ETL tools was a consolidation of 120 separate general ledger systems into a single PeopleSoft instance. Clearly, ETL plays a role in a consolidation project of this size, but so do solutions for analyzing the existing information in the initial stages of the consolidation and improving the quality of that information before any data is loaded into the target application.

Often without explicitly asking for it, companies are really searching for a complete enterprise data integration solution. They want to solve an application upgrade, consolidation or migration problem rather than just purchase an ETL tool. They also require an enterprise data integration architecture, complete with a corresponding implementation methodology, a software platform, IT staff mentoring and best practices and more. This holistic approach will allow them to replace years of hand-coded—that is, inflexible and expensive to maintain—solutions with a “Center of Excellence for Data Integration” (CEDI) specifically for data integration. This CEDI for integrations can be an outstanding vehicle for technology reuse, implementation best practices and mentoring.

Implementing master reference data architecture

For all of these initiatives, harmonizing the master reference data is one of the key initial steps. This activity can become a cornerstone to success and also help fuel any associated enterprise applications such as business intelligence (BI) initiatives, customer data integration, product information management and so on. Therefore, it is imperative to implement a master reference data architecture to help facilitate the consolidation of data, applications and IT infrastructure.

A recent example illustrates this point. A US\$10 billion retailer began its project by evaluating ETL tools that IT believed were needed. However, after deeper discovery with the parties involved, IBM found the company was really undertaking a huge overhaul of its core applications—essentially replacing multiple divisional applications with one enterprise application. It was also planning to implement a single enterprise data warehouse, one payroll system, one financial system and so on. Even its systems integrator soon realized that a systems approach to the project was the only way to succeed in this effort versus the purchase of point solutions and tools.

After a rigorous evaluation of competing ETL tools, the customer and systems integrator chose a software platform approach that would allow them to cover more than just ETL capabilities. Using this approach, they could couple that software architecture with a center of excellence strategy to capture best practices and apply that knowledge to subsequent projects. This strategy would thereby increase the return on investment from the data integration strategy, especially over the long term. Now, they can take what they learned about how to better manage, cleanse and integrate their data, and reapply it as necessary.

Current state versus desired state

A common scenario for companies at the IT crossroads includes:

Current state:

- *Overlapping and redundant data, applications and infrastructure (including servers and storage)*
- *No single, consolidated view of enterprise data*
- *A complex and tightly coupled mix of hand-coded integration programs*
- *Greater than 40 percent of IT budgets committed to supporting this state, which stifles flexibility and ultimately the organization's competitiveness*

Desired future state:

- *Need a radical consolidation of data, applications and infrastructure*
- *Run the business on a single, consolidated view of enterprise data, driven from harmonized master reference data*
- *Untangle and eliminate the hand-coded knot of integration programs by adopting a flexible, reusable and scalable data integration architecture*
- *Reduce costs radically while improving the ability of the organization to respond competitively in its market*

To move an organization from its current state to the future is no easy undertaking. There is a significant bridge to cross between the two—and of course it is not easy. On the other hand, the costs of doing nothing are also huge. In the end, the value to the business can far outweigh the cost of the effort. To stand still and do nothing will likely place the business at risk of failure.

Reduce bottom-line costs, increase top-line growth

IBM customers that have evaluated their current state and the impact of improving their applications and associated IT infrastructure provide some eye-opening examples of the potential benefits of data integration.

- *A major U.S. bank found itself with tens of terabytes of redundant and overlapping data that resulted from a series of acquisitions. By cleaning up the duplicate master reference data and creating a consolidated view of enterprise data, the bank **reduced storage costs by US\$30 million per year** on a single project.*
- *A major global chemicals company was running 12 separate SAP instances and struggling to get a consolidated view of the business. By consolidating to a single global instance of SAP and reducing all the administrative and infrastructure overhead, the company expects to **save operating costs of US\$40 million per year**.*
- *A global telecommunications provider had three different purchase order systems and no way to get a consolidated view of customer orders. By creating an environment that provided a single view of customer orders, the company was able to up-sell and cross-sell its products to **capture an additional US\$200 million in revenue**.*
- *A global logistics provider whose strategy was to grow the business through mergers and acquisitions found itself maintaining 18 data centers, 1,500 applications and 2,600 servers. To address these issues, the company embarked on a huge consolidation project. The target reductions were to maintain only four data centers, 200 applications and 1,600 servers, and to roll out a new enterprise application. This restructuring and consolidation was predicted to **increase profit by more than €1 billion euros**.*

- *A large U.S. bank had 3,000 people hand-coding point-to-point integrations and a Canadian bank had 500 people on the same task. By adopting a data integration platform and eliminating this hand-coding of integrations and overlapping work, both companies projected an **annual 50 percent savings—\$150 million for the U.S. bank and C\$250 million for the Canadian bank.***

This approach to enterprise data integration clearly addresses significant benefits to the companies: reduction of bottom-line costs and increase in top-line growth.

Data integration helps DHL gain market share

The key business issue driving DHL's expansion through mergers and acquisitions in recent years has been its strategic need to provide customers with consistent, superior, price-competitive service on a global basis. DHL clients want information, customer service and service level agreements (SLAs) on a consistent basis for all operations worldwide.

Historically, three DHL global operating groups—North America, Europe and Asia—independently built and operated IT systems to support their region's business. DHL acquisitions and mergers in 2002 and 2003 added additional silos of data. The company developed and designed a key IT program to eliminate these silos by integrating and consolidating its business processes, information systems and customer service functions on a global basis. DHL selected IBM as its enterprise integration platform to power the data integration in this initiative.

An integrated single view of customer records

DHL is using IBM Information On Demand Infrastructure technology to replace manual code development in the movement of financial data from its North American operating systems into its global data warehouse. The IT group at DHL estimates that for this project, *modifications to hand-coded programs would have required nine person-months, but this task now takes less than a week.*

IBM Information On Demand Infrastructure technology allows DHL to meet aggressive timelines set by the company for full integration of the Airborne business. Announced in March 2003 and finalized in the summer of 2003, the DHL acquisition of Airborne's ground operations was designed to help expand DHL's presence in the U.S. and gain market share. This goal was achievable only if DHL could seamlessly integrate and deliver service to Airborne clients. To meet this challenge, IBM helped DHL integrate Airborne's customer data into the DHL systems under an integrated, single view of the customer record.

Additionally, DHL is meeting client demands by providing improved global data on their DHL contracts and service levels. Using the IBM Information On Demand Infrastructure technology and methodology to create a local copy of the company's global data warehouse from Kuala Lumpur, DHL North America can now provide clients with a single integrated view into DHL performance on a global basis. This capability has allowed DHL North America to offer clients information on activity across all major regions—North America, Europe and Asia.

Moving toward an integrated platform

The best approach to achieving the highest business value is to consider a single, truly integrated platform that provides all of the capabilities described above versus buying into multiple stand-alone tools that are not designed to work well together, if at all. Any other approach means cobbling together a series of disparate data integration technologies to make a solution work.

These projects are large, complex and expensive. In addition, most companies are not in the enterprise data integration business, so the required skills generally are not among the core competency within the company. As a result, it is unrealistic to expect companies to take on these integration projects without help from business partners. The most effective and efficient approach is to work with systems integrators and dedicated solution providers.

Getting started

Among the first action items is to harmonize, rationalize and standardize the master reference data within the company. Many business initiatives can benefit from this accurate and reliable information. In fact, that is one reason it is a hot topic of discussion for enterprise application vendors.

The issue around “messy” master data is its inconsistency across the enterprise. For example, different data values uniquely describe a business entity used to distinguish one item from another, such as customer name, address, date of birth, part number, item code and so on. Different identifiers are assigned to each unique instance of a business entity. Different relationships between business entities—for example, two customers—are “householded” together at the same location or address. Different hierarchies exist among business entities; for example, a parent company owns other companies. All of these facts across the heterogeneous IT systems landscape conspire against the ability to get answers to simple questions such as: “How many customers do we have?”, “Who are the most profitable customers?”, “How much do these parts cost from our different suppliers?” and “Which products have these customers purchased?”

The obvious questions for the organization include:

- *How do we overcome the “messy” data issue?*
- *How do we design and implement an enterprise data architecture to avoid this over the long term?*
- *What methodology can help to harmonize this master data?*
- *How do we reduce risk, cost and timelines to accomplish this?*
- *What technology can help automate manual processes, integrate data integration processes, write business rules once and reuse them, and exploit data across applications?*
- *How do we organize the stakeholders, use the lessons learned, develop a center of excellence and then apply these lessons to the rest of the project?*

Using an enterprise data integration platform

From the IBM view, an integrated software platform, illustrated in Figure 1, offers many benefits:

- **Connects data sources with targets** and can support a wide variety of sources—*from mainframe, native adapters and other legacy systems as well as major ERP software including SAP, Siebel, Oracle, PeopleSoft Enterprise and PeopleSoft EnterpriseOne (formerly JD Edwards OneWorld)*
- **Helps companies profile and understand the structure and content of heterogeneous data** in terms of *what state the data is in and the relationships of the data*
- **Identifies data that needs to be cleansed, standardized and harmonized**—*whether it is product, customer, supplier, or names and addresses—and this also includes global address validation*
- **Transforms data** to map the source to the target, *enriches it with external data and then connects and moves it to the target; depending on the scenario, this might be to the same vendor solution or to a different one*
- **Manages metadata**, *not only the metadata generated throughout the process, but also a view to the metadata generated by other tools, such as data modeling tools*

- *Can handle very large volumes of data in parallel—enabling large loads to be completed in batch—but also exposes these data harmonization routines as Web services to be called on demand; integrity of the core data can be easily maintained when new customer records or item records are added to the system*
- *Leverages an implementation methodology to help get the most from the technology much more quickly and helps form the basis of a Center of Excellence for Data Integration*

Figure 1: Enterprise data integration platform from IBM

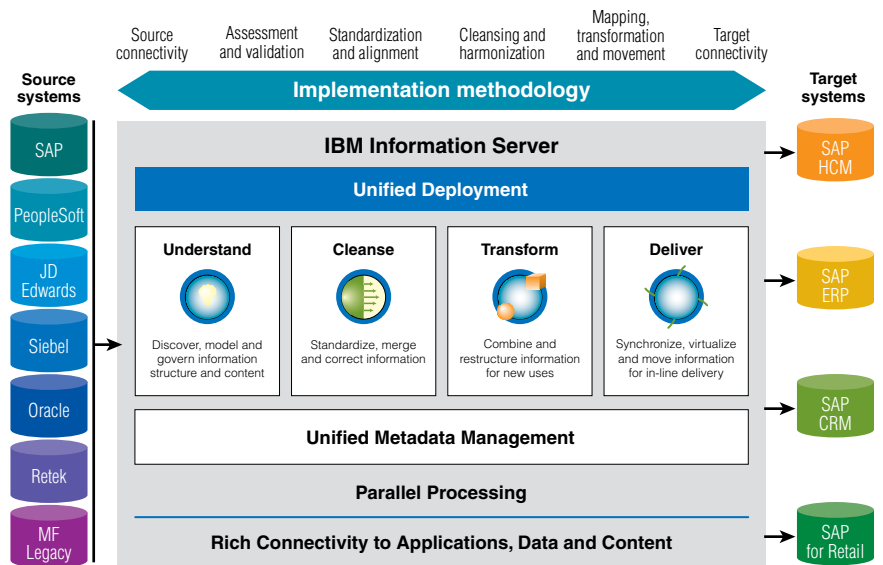


Figure 2: An SAP example of a platform approach to harmonize master data

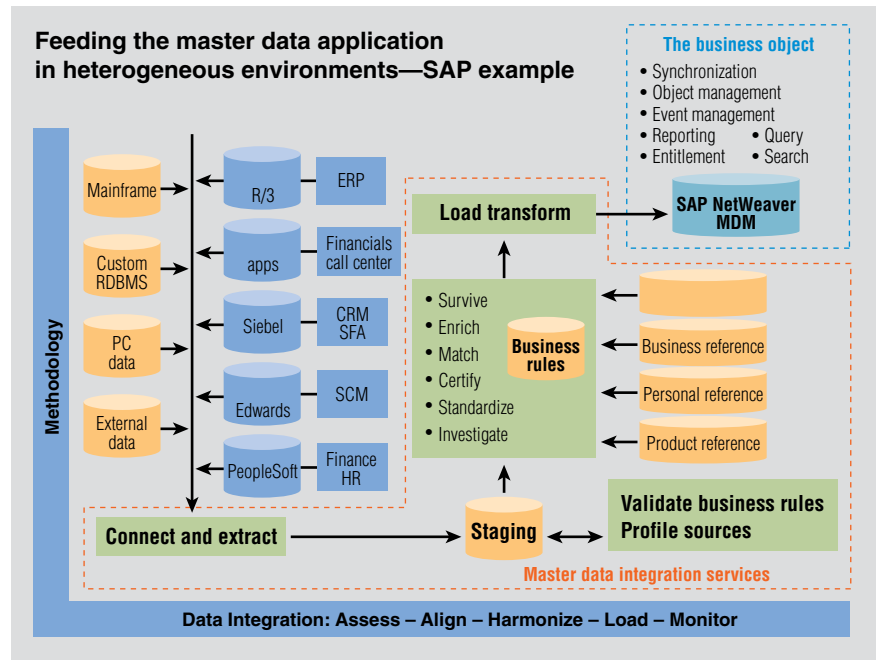


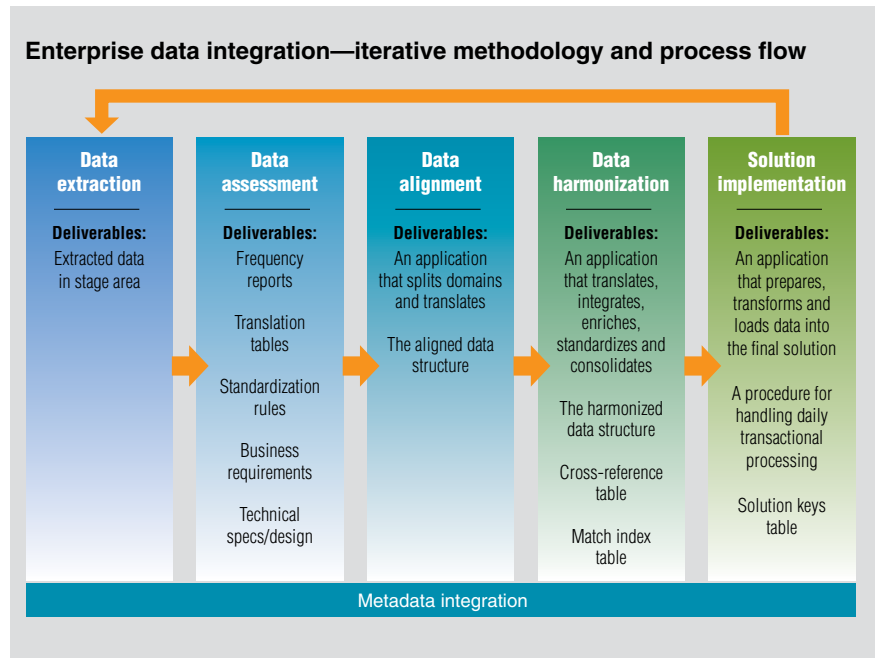
Figure 2 shows an example of using this platform approach to harmonize master data. It represents a large heterogeneous environment with many disparate sources handling different aspects of the business. The master data integration service (shown encircled in orange dashed lines) provides the ability to connect to these heterogeneous sources, allowing validation of business rules and profiling of the sources. The data quality process involves investigating the data, finding the duplicates, developing standardization rules, certifying data against a global address standard when that is the case,

matching the data, enriching it with external reference data (such as Dun & Bradstreet information) and creating a surviving record. This record can then be mapped and subsequently loaded into the master data application, in this case the SAP NetWeaver® Master Data Management (SAP NetWeaver MDM) components. The master data becomes a “business object” that SAP NetWeaver MDM then “owns” and controls for entitlement, synchronization, reporting, searching, querying and other capabilities.

The value provided by IBM in this scenario is the ability to leverage a broad range of capabilities provided by IBM® Information Server for accessing, profiling, cleansing, transforming and moving data so that an application such as SAP NetWeaver MDM can be deployed with clean data from the outset. It can also be done more quickly than if the point of convergence and data cleansing were to take place inside this application. IBM tools and methodology are designed to address this process and make it as fast and efficient as possible in order to deliver maximum return on investment (ROI) to the customer.

Figure 2 also shows SAP as one of the heterogeneous sources. Since SAP has its own connectors, customers need to make an architectural decision whether to load the SAP operational data directly into SAP NetWeaver MDM and then load all the non-SAP data and merge it inside SAP NetWeaver MDM, or whether to merge all SAP and non-SAP data once and then do a single load into the SAP NetWeaver MDM component. This decision can be made based on individual circumstances and issues such as data transformation and mapping requirements, data volumes involved, whether the business rules remain the same or need to be redefined and so on.

Figure 3: Process flow using IBM methodology for migration, consolidation and upgrades



Implementation methodology to create a scalable, repeatable solution

Figure 3 shows a view of the process flow associated with the IBM methodology for implementing application migrations, consolidations and upgrades with harmonized master data. This approach is an iterative process from both end to end and within each of the five phases. Each phase corresponds to the items listed in the solution diagram in Figure 1. This methodology, coupled with additional best practices, forms the basis of a Center of Excellence for Data Integration that many companies are adopting.

The benefit is twofold: a complete, scalable solution that can be applied to these data harmonization challenges and a methodology that not only includes best practices but is also repeatable and can be applied to subsequent projects. A CEDI enables data integration costs to be reduced, especially over the long term.

By incorporating the knowledge learned from initial projects, IBM customers are realizing tremendous benefits and savings on subsequent integration projects. The combined solution of technology and methodology can successfully help move from the current state to the desired state during enterprise application migrations, upgrades and consolidations in large heterogeneous systems landscapes. The CEDI approach is helping many customers complete projects on or before schedule and thus, under budget.

EDF Energy dramatically reduces costs using data integration methodology

EDF Energy was formed through the merger of two equal-sized companies: SeeBoard and London Electric. EDF Energy now supports over 25 percent of the electricity needs in the United Kingdom. However, as a result of the merger, EDF Energy had two, and sometimes more, of each application required to run a regulated utility. This was clearly inefficient and costly.

To compound these issues, one company ran all its systems in house while the other outsourced everything to a third party. One goal of the merger was to consolidate systems to help reduce costs and to develop a consolidated view of the customer to facilitate up-selling and cross-selling of products and services. Additionally, operational systems could not experience downtime during the conversion because the company needed continuous usage monitoring and billing capabilities to maintain its revenue stream.

To meet the tight project deadline, a systems integrator proposed hand-coding that was projected to consume 4,000 person-days of work. The math quickly showed that this approach would not allow EDF Energy to make its switchover deadline. But the high cost of not switching numbered in the millions of pounds Sterling by choosing this hand-coded approach.

IBM scoped the project by automating as much as possible using technology from the IBM Information On Demand Infrastructure team and employing the IBM® Iterations® methodology also from this infrastructure. This approach yielded a **90 percent labor reduction**—a 400 person-day effort compared to the original 4,000 day estimate. Using this approach, EDF Energy **finished the project four months early and estimated a savings of £16 million.**

IBM Information On Demand Infrastructure solutions helped EDF Energy profile, standardize and transform the data and business rules in this migration and consolidation project. EDF Energy was able to consolidate the systems, bring them in house and meet its project deadline.

Seven strategies to help reduce risk and costs

To summarize, some data integration lessons learned and strategies to undertake during application migration or consolidation include:

- 1. Use an automated source system profiling technology to assess the structure, quality and content of source system data. The right technology can also assist with business rule identification and validation.*
- 2. Allow for connectivity across the enterprise—to mainframes, UNIX® and Microsoft® Windows® systems, custom-built applications and third-party software provided by major enterprise application vendors and others.*

- 3. Improve data quality** as part of the migration project by standardizing data attributes, correcting data quality issues with names, parts, addresses, suppliers and other business entities, and matching data to identify and eliminate duplicate data records and identifiers.
- 4. Reconcile conflicting application data models** by assessing technical source data and identifying and validating business rules, which enables easier mapping and re-engineering of data to a defined target system.
- 5. Help ensure that very large data volumes can be handled** by using a solution that supports parallel processing capabilities, so scalability increases linearly as additional hardware resources are added.
- 6. Manage metadata throughout the project by collecting business, technical and operational metadata.** The right tools can provide a view to data lineage and impact analysis to help evaluate the nature of changes to the data as well as where and how these changes affect other parts of the process.
- 7. Define and create a data integration “center of excellence”** with a complete methodology that includes defined roles and responsibilities for the staff involved and a repeatable process to understand, consolidate, cleanse and integrate data.

For more information

For more information, please contact your local IBM Information On Demand Infrastructure representative or visit ibm.com/software/data/integration



© Copyright IBM Corporation 2007

IBM Software Group
Route 100
Somers, NY 10589

Printed in the United States of America
March 2007
All Rights Reserved

IBM, the IBM logo and Iterations are trademarks or registered trademarks of International Business Machines Corporation in the United States, other countries or both.

Microsoft and Windows are trademarks of Microsoft Corporation in the United States, other countries or both.

UNIX is a registered trademark of The Open Group in the United States and other countries.

Other company, product or service names may be trademarks or service marks of others.

References in this publication to IBM products or services do not imply that IBM intends to make them available in all countries in which IBM operates. Offerings are subject to change, extension or withdrawal without notice.

All statements regarding IBM future direction or intent are subject to change or withdrawal without notice and represent goals and objectives only.

The information contained in this document is provided for informational purposes only. While efforts were made to verify the completeness and accuracy of the information contained in this document, it is provided "as is" without warranty of any kind, express or implied. In addition, this information is based on IBM's current product plans and strategy, which are subject to change by IBM without notice. IBM shall not be responsible for any damages arising out of the use of, or otherwise related to, this document or any other documents. Nothing contained in this document is intended to, nor shall have the effect of, creating any warranties or representations from IBM Software.

TAKE BACK CONTROL WITH **Information Management**