**Information integration
challenges and benefits
for enterprise application
migrations, consolidations
and upgrades.**

Contents

Introduction

Many companies today are at an information technology (IT) crossroads, and for a variety of reasons. First, their IT departments are being asked to deliver more, but with less. There are a number of key business initiatives that cannot succeed without major IT intervention – initiatives such as single view of a customer, strategic sourcing, enhanced inventory management, and preparing for regulatory compliance for Sarbanes-Oxley (SOX) and other governmental reporting entities. These initiatives all require consolidation flexibility around the data.

Making this difficult is the fact that there usually are many enterprise application instances in an organization – applications such as ERP, CRM, SCM and SRM from vendors such as SAP, Oracle, PeopleSoft, JD Edwards and Siebel. Additionally, a large number of 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 aggravated by corporate mergers and acquisitions where the resulting entity ended up with two – or more – of every application.

Harnessing and making sense of this data is what is getting in the way of being able to feed these key business initiatives. One issue is inconsistent master reference data. It is a huge problem, and it is becoming more talked about today than ever before. Companies are therefore pursuing application migration and consolidation strategies to have a positive impact on enabling top-line growth strategies.

In addition to the top-line growth benefits an enterprise application migration and/or consolidation can afford, this complex application landscape situation affects the bottom line. All these applications are expensive – costs are multiplied by the number of instances, software license costs, software maintenance renewal costs, hardware maintenance and upgrade costs, and staffing costs.

The other factor compounding this crossroads is the recent enterprise application vendor consolidation. For example, JD Edwards and PeopleSoft customers are faced with at least reviewing the decision of what to do with their current enterprise application – do I stay with my current vendor and upgrade it? Do I switch vendors? If I do that, how will I quickly re-implement my ERP? How will I transform all my business rules and data? Or maybe I stay with what I have but I consolidate other systems to address the disparate data issues that are keeping me from having harmonized master reference data that delivers my single view of a customer and powers other key business initiatives.

Data integration issues and challenges

Whatever the case, there is really no getting around the problem with data access and data quality whether you are consolidating applications, or eventually upgrading them, or migrating to another vendor, or simply staying at the status quo and just getting your arms around the data so you can deliver a 360 degree view of a customer or a supplier, or prepare for SOX and other regulatory reporting burdens.

These are the data integration issues you will face:

- *The enterprise application source meta data is not easily assembled in one spot to understand what is available. There may also be legacy sources added to the mix, which often don't make any meta data available through a standard API, if at all.*
- *The "master reference data" – names and addresses for suppliers and customers, the part numbers and descriptions – is not the same, and/or there are duplicate sources of this data maintained between instances of a single application and between different ERP, CRM and SRM applications.*
- *Hundreds of extraction, transformation & loading (ETL) jobs need to be written to move data from all the sources to the new target application.*
- *Data transformations will be required prior to load to fit the data into the new environment structures.*

- *Making sure this work can be easily re-run, and maintained, by new staff, as IT brings different teams on line or as the roll out occurs in phases.*

- *Ensuring you can handle large amounts of data can be run through the process, and that it will finish on time; and the contrary: can the infrastructure support running any of the above transformations and data matching on demand?*

As IBM has discussed these issues and challenges with customers and systems integrators who frequently are contracted to do the work, we've found that at the outset a lot of people do not really know how to go about getting started. They may have heard of some tools to 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 case we see is companies who think they just need a single tool, whether it is an enterprise application integration (EAI) tool, or an extract, transform and load (ETL) tool, or they think they just need to write a few custom programs in a simple, well-known language.

What is really going on is this is a much more complex project than just extracting, transforming and loading data. For example, one customer who was shopping for an ETL tool finally disclosed what they were really trying to accomplish was a consolidation of 120 separate general ledger systems into PeopleSoft.

What companies are really asking for, often without explicitly asking for it, is a complete data integration solution. They want to solve an application consolidation or migration or upgrade problem rather than just buy an ETL tool. And they also require an enterprise data integration architecture, complete with a methodology, a software platform, IT staff mentoring and best practices. This will allow them to replace years of hand-coded solutions with a data integration "Center of Excellence." The center of excellence can be the vehicle for technology reuse, implementation best practices and mentoring.

In any of these initiatives, getting the master reference data harmonized is one of the key steps. This activity can be a cornerstone to success and will also help fuel any associated business intelligence (BI) initiatives driven off the new, harmonized and rationalized enterprise application landscape. It is therefore imperative to implement a master reference data architecture and business application to help facilitate consolidating your data, your applications, and IT infrastructure.

To illustrate this point, here is a quick example. A US$10 billion retailer started out shopping for an ETL tool, when in fact, after discussing with the parties involved, they were doing a huge overhaul of their core applications – replacing divisional applications with one enterprise application. They were also planning on one data warehouse, one payroll system, one financial system and so on. Even their systems integrator (SI) soon realized that the only way to succeed was to take a systems approach.

After a rigorous evaluation of competing ETL tools, the customer and SI soon realized the only way to tackle it was with a software platform approach that would allow them to cover more than just ETL capabilities, and to couple that software architecture with a center of excellence strategy to be able to reuse everything they would learn in subsequent projects. This strategy would thereby increase their return on investment from their data integration strategy. Now, they are able to take what they learned about how to better manage, cleanse and integrate their data, and reapply it as necessary.

Current state vs. desired state: Why bother?

Here is a common scenario for companies at this IT crossroads:

Current state:

- *Overlapping and redundant data, applications and infrastructure (including servers and storage)*
- *No single, consolidated view of their enterprise data*
- *A hairball mix of hand-coded integration programs*
- *Greater than 40% of the IT budget is committed to supporting this state, which chokes flexibility and ultimately the organization's competitiveness.*

Desired future state:

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

This all sounds great. However, there is a big gap to bridge between the current and the future states, and of course it is not easy, and it is a big project. But the costs of doing nothing are huge, and the business returns can be even larger.

Here are some examples from IBM customers that compare the current issue – or opportunity, depending on your outlook – and what the impact can be:

- *A major US bank had tens of terabytes of redundant and overlapping data following a series of acquisitions. By cleaning up the duplicate master reference data and creating a consolidated view of enterprise data, the bank reduced storage by US$30 million per year on a single project.*

- *A major global chemicals company is running twelve SAP instances and struggles 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 will **save operating costs of US$40 million per year.***
- *A global telecommunications provider had three order systems and no way to get a consolidated view of customer orders. By creating an environment to get 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 who grew by merger found themselves maintaining 18 data centers, 1500 applications and 2600 servers. They embarked on a huge consolidation project reduce down to four data centers, 200 applications and 1600 servers, and rollout a new enterprise application. This restructuring and consolidation is predicted to **increase profit by over €1 billion euros** by the end of 2005.*
- *A large US bank had 3000 people hand coding integration programs and a Canadian bank had 500 people on the same task. By adopting a data integration platform, getting rid of hand-coded applications and reducing resources, both projected an **annual 50% savings for US$150 million and C$250 million,** respectively.*

These benefits clearly address both the bottom line costs and contribute to top line growth.

What should I do first?

The right way to think about this and approach this is not to pigeon-hole a solution from a friendly acronym you may know – EAI, ETL, BPM, EII et al – but to fundamentally think about merging these integration needs from one integrated platform. Anything else means cobbling together a series of disparate data integration technologies to make a solution.

And these projects are big enough as it is. Since most companies are not in the enterprise data integration business, it is not among their core competencies, so it is too much to expect companies to be able to figure this out on their own. Systems integrators and dedicated solution providers are the right approach.

Among the first steps is to harmonize, rationalize and standardize your master reference data, because many business initiative solutions can be born from this activity. This is why the enterprise application vendors are talking about it so much lately.

The issue around messy master data comes from that fact that it is generally inconsistent across the enterprise. There are different data values that uniquely describe a business entity used to tell one from another, such as customer name, address, date of birth, part number, item code, etc. There are different identifiers assigned to each unique instance of a business entity. There are different relationships between business entities – two customers "householded" together at the same location or address. There are different hierarchies among business entities – for example, a parent company owns other companies. All these facts across your heterogeneous IT systems landscape conspire against the ability to come up with 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 then are, how do you overcome this? How do you design and implement an enterprise data architecture? What is the methodology to harmonize your master data? How do you reduce risk, cost and timelines to accomplish this? What technology should you use to automate manual processes, integrate data integration processes, write business rules once and reuse them, and exploit data across applications? How to you organize the stakeholders and how do you take the lessons learned and develop a center of excellence to apply these lessons to the rest of the project?
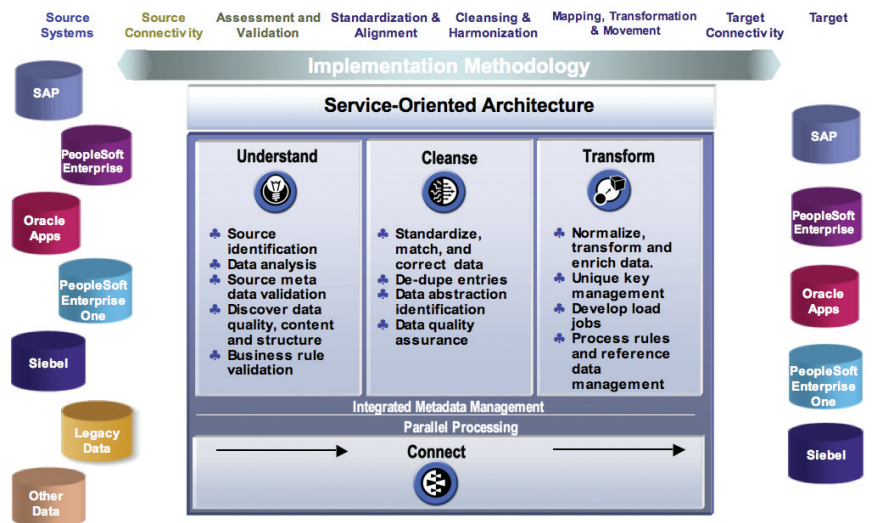
Figure 1. Rapid Deployment Framework.

Using an enterprise data integration platform

IBM's view is you can greatly benefit from an integrated software platform (Figure 1) that can:

- *connect to your sources and targets – to support a wide variety of sources, from mainframe, native adapters and other legacy systems, as well as the major ERP players including SAP, Siebel, Oracle, PeopleSoft Enterprise and PeopleSoft EnterpriseOne (formerly JD Edwards OneWorld);*
- *help you profile and understand the structure and content of your heterogeneous data so you know what state it is in and how it all relates to one another;*
- *identify what needs to be cleansed, standardized and harmonized, on any kind of data, whether it is product, customer, supplier, or names and addresses, including global address validation;*
- *transform data to map source to target, and enrich it with external data, and then connect and move it to the target. Depending on the scenario, this might be to the same vendor solution or to a different one.*
- *manage meta data, not only the meta data generated throughout the process, but also a view to the meta data that other tools, such as ER tools, generate;*

- *be able to handle not only very large amounts of data in parallel – so you can complete large loads in batch - but also exposing these data harmonization routines as web services to be called on demand. This will allow you to maintain the integrity of the core data when new customer records or item records are added to the system.*
- *leverage an implementation methodology to help you get the most out of the technology, quicker, that can form the basis of a data integration center of excellence.*

Here is an example of using this platform approach to harmonize master data. In figure 2 is a large heterogeneous environment with many disparate sources handling different aspects of the business.

The area encircled in orange dashed lines is the master data integration service that includes the ability to connect to these heterogeneous sources so you can validate the business rules and profile the sources, then run through the data quality process of 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 as necessary), and creating a surviving record, that can then be mapped and subsequently loaded into the master data application, in this case SAP MDM. The master data becomes a "business object" that SAP MDM then "owns" and controls for entitlement, synchronization, reporting, searching, querying and other capabilities. The value that IBM is providing in this scenario is an ability to leverage the broad range of capabilities the IBM WebSphere® Data Integration Suite provides for accessing, profiling, fixing and moving data, so that an application such as SAP MDM can be deployed with clean data from the outset, and done more quickly than if the point of convergence and data cleansing were to take place inside this application. IBM's tools and methodology are designed to address this process and make it as fast and efficient as possible.
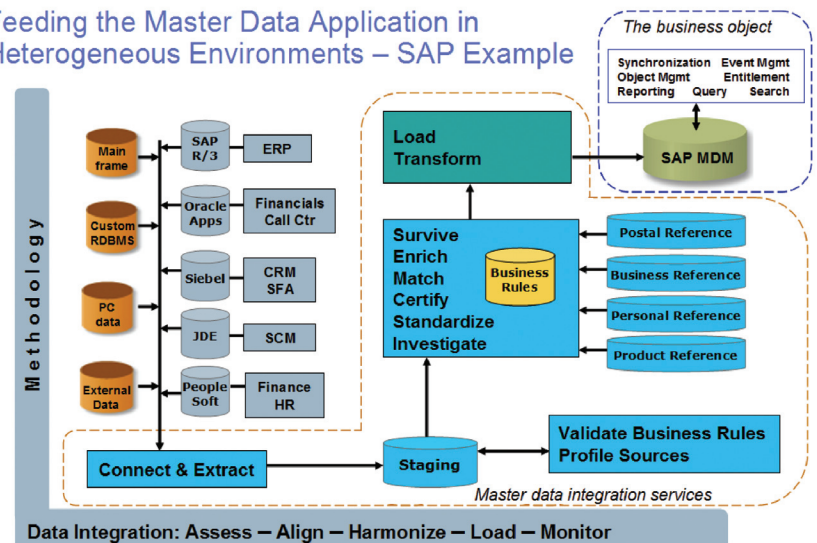
Figure 2. Master data application in heterogeneous Environments – SAP example.

Figure 2 also shows SAP as one of the heterogeneous sources. SAP has its own connectors to itself, so one architectural decision customers need to make is whether to load the SAP operational data directly into SAP MDM, and then load all the non-SAP data and merge it inside SAP MDM, or whether to merge all SAP and non-SAP data once and then do a single load into the MDM application. This is a decision that can be made based on your individual circumstances and how much data transformation and mapping may be required, the data volumes involved, whether the business rules are relatively the same or need re-definition, and other factors.

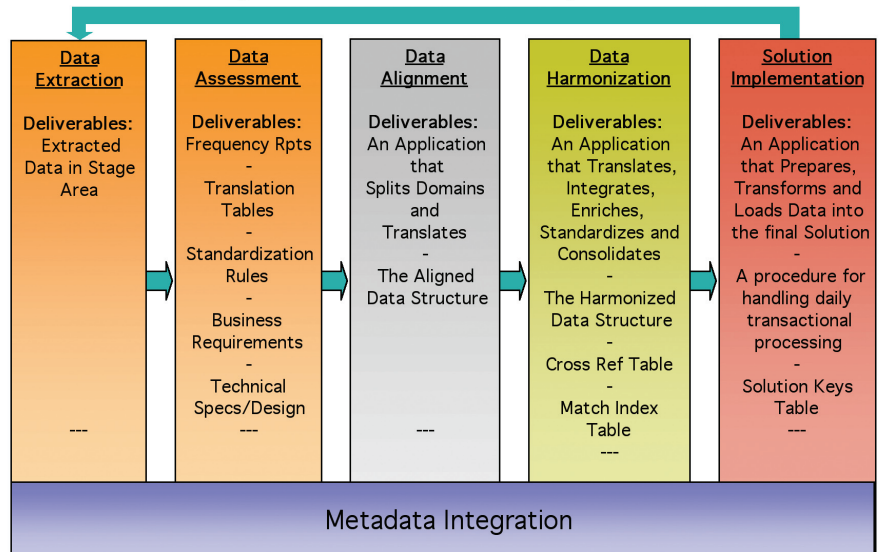### Enterprise Data Integration–Iterative Methodology & Process Flow

Data Extraction	Data Assessment	Data Alignment	Data Harmonization	Solution Implementation
Deliverables: Extracted Data in Stage Area	**Deliverables:** Frequency Rpts - Translation Tables - Standardization Rules - Business Requirements - Technical Specs/Design	**Deliverables:** An Application that Splits Domains and Translates - The Aligned Data Structure	**Deliverables:** An Application that Translates, Integrates, Enriches, Standardizes and Consolidates - The Harmonized Data Structure - Cross Ref Table - Match Index Table	**Deliverables:** An Application that Prepares, Transforms and Loads Data into the final Solution - A procedure for handling daily transactional processing - Solution Keys Table
---	---	---	---	---

Metadata Integration

Figure 3. Iterative methodology & process flow.

Implementation methodology

Figure 3 shows a view of the process flow associated with IBM's methodology for implementing application migrations, consolidations and upgrades with harmonized master data. This is an iterative process both from end to end, and also within each of the five phases. Each of these phases corresponds to the items listed in the solution diagram in Figure 1. This methodology plus additional best practices forms the basis of a data integration center of excellence that more and more of companies are adopting. The benefit of this is not only do you have a complete and scalable solution to apply to these data harmonization challenges, but you also have a methodology that is repeatable, includes best practices and, coupled with IBM services, mentoring, and can therefore be applied to subsequent projects. A Center of Excellence lets you reduce data integration costs. IBM customers have found that by incorporating the learnings and best practices from initial projects, later projects are done more quickly and more work can be reused. This has proven to lead to faster time-to-benefits from the resulting application. The complete solution of technology and methodology can help you move from the current state to the desired state during

enterprise application migrations, upgrades and consolidations in large heterogeneous systems landscapes. The Center of Excellence approach will help you become self-sufficient for future initiatives, and our customers have been able to deliver these often early and under budget.

Customer case study: DHL Corporation

The key business issue driving DHL's expansion through mergers and acquisitions in recent years has been their strategic need to provide their customers with consistent, superior, price competitive service on a global basis. DHL's clients want information, customer service and service level agreements on a consistent basis for all operations worldwide. Historically DHL's three global operating groups – North America, Europe and Asia – independently built and operated IT systems to support their region's business. DHL's acquisitions and mergers in 2002 and 2003 added additional silos of data. A key IT program was developed and designed to eliminate these silos by integrating and consolidating their business processes, information systems, and customer service functions on a global basis. DHL selected IBM as their enterprise integration platform to power the data integration in this initiative.

DHL is using IBM's technology to replace manual code development in movement of financial data from their North American operating systems into their global data warehouse. DHL's IT group estimates that for this project, modifications to hand coded programs would have taken 9 person-months; this task now takes less than a week. DHL's use of IBM's technology is allowing them to meet the aggressive timeline the company set for full integration of the Airborne business. Announced in March 2003 and finalized in the summer of 2003, DHL's acquisition of Airborne's ground operations was designed to help expand DHL's presence in the US and gain marketshare. This goal was only achievable if DHL could seamlessly integrate and deliver service to Airborne's clients. IBM helped DHL meet this challenge by allowing them to integrate Airborne's customer data into their systems, under an integrated, single view of the customer record. Additionally, DHL is meeting client demands for better global data on their DHL contracts and service levels. Using IBM's technology and methodology

to create a local copy of the company's global data warehouse from Kuala Lumpur, DHL North America is now able to provide clients with a single integrated view into DHL's 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.

Customer case study: EDF Energy

EDF Energy was formed through the merger of two equal sized companies, SeeBoard and London Electric. Combined, EDF Energy now supports over 25% of the UK population's electricity needs. However, once the organizations were merged, EDF Energy ended up with two, and sometimes more, of each application required to run a regulated utility. Compounding this was the fact that one company ran all their systems in house while the other outsourced everything to a systems integrator. One of the goals of the merger was to be able to consolidate systems to reduce costs and develop a consolidated view of the customer to facilitate up-selling and cross-selling of products and services. Additionally, operational systems needed to be converted with no downtime to continue usage monitoring and billing capabilities to maintain their 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 their switchover deadline. They calculated a very high cost of not switching that numbered in the millions of pounds Sterling by taking this hand-coded approach. IBM scoped the project by automating as much as possible using technology in the IBM WebSphere Data Integration Suite and employing the IBM Iterations® for IBM WebSphere Data Integration Suite methodology. This yielded a 90% labor reduction – a 400 person-day effort compared to the original 4,000 day estimate. EDF Energy took this approach, finished the project 4 months early and estimated a savings of £16 million.

EDF Energy used IBM Information Integration Solutions to help profile, standardize and transform the data and business rules in this migration and consolidation project. EDF Energy were able to consolidate the systems, bring them in house and meet their project deadline.

Consolidation vs. federation

So far, this paper has discussed the case for consolidation. It is a strong business case because it can affect both bottom-line and top-line goals. However, it is a large project and can take a long time to complete. Consolidation is a more traditional data integration process. The data is treated primarily in bulk, and a big assumption is the data is primarily structured data – think of record-oriented data as an example.

But a lot of data might be unstructured. It might be other kinds of content – images, for example, with many attributes but no "record-oriented" structure. In certain applications this kind of data is important. Further, some types of data, whether it is structured or unstructured, may not belong to your organization because it is held in an external system or application. There may be licensing restrictions or other controls on that data that prevent you from consolidating it in a new system. The target for this data may be a user query & analysis tool rather than a persistent store in a database. The use case typically is more on-demand than bulk-oriented.

For these reasons, even if you have primarily structured data, you may want to consider a federation approach as part of your consolidation project. While attempting to integrate organizations after a merger, public companies are under pressure to show results quickly, and getting at customer cross-sell and upsell opportunities may not be able to wait until you can merge together two disparate IT heritages. Federation can give you a headstart on leveraging your data in this scenario.
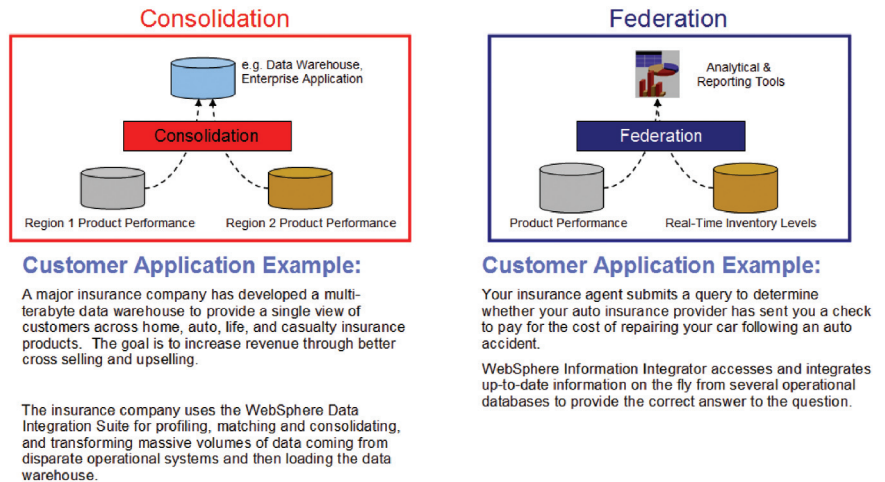
Consolidation Versus Federation

Figure 4. Consolidation and federation by use case example.

Figure 4 shows two examples of the types of applications, and you can see the subtle difference. In consolidation the company is setting up for ongoing management of the business by merging systems to get a single view of customer across lines of business. In the federation example, it is more operational in nature and therefore more appropriate for an ad hoc query shown here. While the consolidation is underway, the federation approach lets a service agent answer a customer issue right away without having to wait for a long consolidation project to complete.

Federation is a way to provide a unified view across a broad range of sources, and so the users don't need to know where the data sources are and how they are structured.
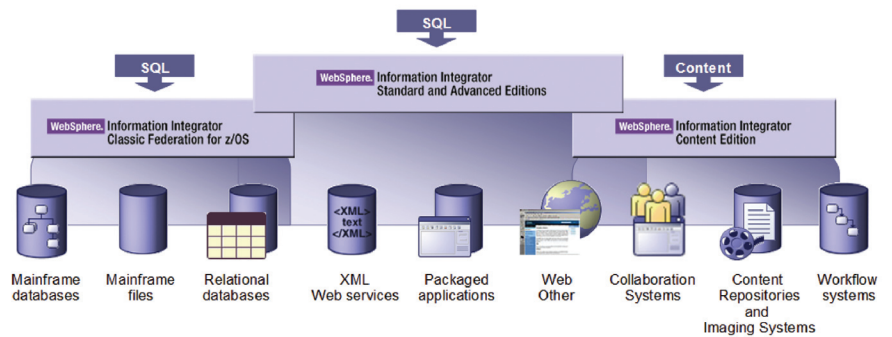
Figure 5. About IBM WebSphere Information Integrator.

WebSphere Information Integrator (see Figure 5) can provide federated information as a web service or via a Java API, and supports both query and update capabilities. It can optimize distributed queries across its backend sources. These sources include mainframe systems, enterprise sources, and content-based systems, repositories and workflow systems. The different editions of WebSphere Information Integrator − Classic Federation, Standard Edition and Advanced Edition, and Content Edition− can be used singly or together to create a rich solution that leverages the data sources that each support.

Figure 6 shows some heuristics on when to federate, when to consolidate, and when to consider both. Starting with common functions and reading clockwise, you can see some of the characteristics of the data and the use cases you should consider in deciding what approach will work best for you.

When to Use Which Approach

Common Functions	Federate
▪ Combining data from various sources ▪ Transforming, cleansing, validating data ▪ Making data available to SQL applications	▪ Unstructured, semi-structured data ▪ Need to access source directly ▪ Licensing restriction on copying ▪ Not all sources can be consolidated ▪ Choice of data sources is dynamic
Consolidate AND Federate	**Consolidate**
▪ Combine real-time data with historical data ▪ Combine unstructured data with historical structured data ▪ Federate for agility during consolidation process	▪ Supports rich transformations ▪ Reduce infrastructure resources, systems and costs ▪ Performance and scalability of data access ▪ Need persistent data storage

Figure 6. When to consolidate and when to federate.

Eight strategies to reduce risk and cost

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

1. *Eliminate source system mystery by using 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 your enterprise, including to mainframes, UNIX & Windows systems, custom-built applications and third party software provided by major enterprise application vendors and others.*

3. *Improve your data quality as part of your migration project, by standardizing data attributes, correcting data quality issues with names, parts, addresses, suppliers and other business entities, and data matching to identify and eliminate duplicate data records and identifiers.*

4. *Reconcile your conflicting application data models through technical source data assessment and business rule identification and validation, which enable easier mapping and re-engineering of data to your defined target system.*

5. *Ensure you can handle very large data volumes by employing a solution that supports parallel processing capabilities, so scalability of the solution increases linearly as you add hardware resources.*

6. *Manage your meta data throughout the project by collecting business, technical and operational Meta Data. Using the right tools, you can get a view to data lineage and impact analysis to evaluate the nature of changes to the data and where and how these changes affect other parts of the process.*

7. *Define and create a 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.*

8. *Balance the need for both consolidation and federation. Examine your data sources and think about who is going to use them and how, to help you prepare a strategy.*

To learn more

Please visit **ibm.com** or contact your local IBM representative.

IBM