A discussion of information integration solutions
November 2005

WebSphere® software

Deploying a Center of Excellence for data integration.

Contents

1 Summary
1 Introduction
2 Mastering data integration
2 The Center of Excellence
3 Focus on people, process, and technology
4 IBM's Center of Excellence for data integration
7 Deploying IBM's Center of Excellence for data integration
8 Center Of Excellence workshops
11 Optional elements
13 Handbooks, implementation guides, plans, and training
15 The IBM WebSphere Data Integration Suite
18 Summary
18 Next steps

Summary

This paper describes:

- *The business case for creating a Center of Excellence for data integration*
- *The IBM Center of Excellence for data integration offering (COE)*
- *How to establish an enterprise data integration Center of Excellence within your company*
- *How the COE can help speed adoption, improve knowledge sharing, and lower the total cost for an enterprise data integration initiative.*
- *A detailed look at the Intellectual Property and content used in implementing the IBM COE offering*
- *How a COE can broaden the reach and context of enterprise application deployment, migrations, consolidations, and upgrades.*

Introduction

There is no shortage of business issues that are leading organizations to construct and leverage a complete, reliable, enterprise-wide view of data. Information Technology groups are creating new systems and adapting old systems in order to:

- *Establish a single view of customers*
- *Optimize supply chains and distribution channels*
- *Rationalize information from mergers and acquisitions*
- *Manage risk and comply with regulatory requirements*
- *Reduce IT costs related to redundant data and systems*

Many of these business initiatives are difficult to achieve because they involve very complex data integration processes. A recent survey conducted by IBM found that:

- *Sixty percent of CEOs believe their organizations need to do a better job of capturing and understanding information in order to make swift decisions*
- *Sixty six percent of CFOs believe that information is not integrated in a cost effective way*
- *Twenty five percent of the companies reported that they have more than 15 different repositories*

These results are not unique. Industry research shows that it is not unusual for a billion dollar company to have 48 different financial systems and 2 to 3 different ERP systems. Gartner analysts have identified clients that utilize 26 different business intelligence technologies and research from Forrester found that as much as 40% of an IT project budget is spent on integration.

Mastering data integration

Many companies allocate funds for data integration technology and services on a project basis. From a budgeting perspective this makes sense, but it leads to a duplication of effort as each project team develops their own set of transformations, using whatever tools or resources they feel are appropriate. All too often there is no leverage from one project to the next.

Over time, the number of data integration techniques, technologies and processes supported by IT, impacts available resources that can be applied to new projects. Data integration processes that are hand coded or require significant customizations add a particular burden for support on the organization, making each new project more difficult and expensive.

Enterprise data integration requires domain expertise. Owning appropriate technology and having development staff trained to use it is just a beginning. To create consistent cross enterprise views of information, organizations must adopt a standard approach or methodology to data integration that reinforces the sharing of best practices and application of common business rules.

The Center of Excellence

In order to meet these challenges, organizations are adopting a centralized approach to managing projects by creating a Center of Excellence for data integration (COE). The COE is staffed by people with broad data integration expertise, who can act as a central point of contact to facilitate collaboration between the lines of business, enterprise architects, database administrators and application developers.

The COE provides governance and direction and may provide the resources to staff each of these components:

- *Architecture*
- *Methodology*
- *Best practices and standards*
- *Education programs*
- *Support services*

Bringing together a dedicated team, with the appropriate domain expertise, enables these companies to:

- *Identify core patterns of integration*
- *Share best practices*
- *Leverage data integration skills and standard tools across the enterprise*

Focus on people, process, and technology

There are three keys components to a successful COE. The COE needs to be staffed with people who bring domain expertise about the business and data integration. The staff needs to have a standard methodology and best practices that will bring consistency and leverage to development projects. The COE needs a technology platform for enterprise data integration that will support the team in all phases of a data integration project including discovery, analysis and mapping, cleansing and standardization, transformation and harmonization, and delivery and synchronization.

Developing domain experts enables these organizations to identify the reference architectures and best practices to guide data integration project teams. Sharing best practices and expertise promotes the reuse of common business rules and data transformations that improve the quality of information and business processes.

Promoting the use of a standard set of data integration tools enables developers to be more productive, reduces the total cost of maintaining integration processes, and provides the organization with a better way to manage and leverage business, technical and project metadata.

By providing project plans with defined roles and responsibilities and skill requirements, the COE enables project teams to benefit from the experience and knowledge gained from every other data integration project. This greatly reduces the risks associated with each additional project and helps minimize the time to value for each project.

IBM's Center of Excellence for data integration

Creating a repeatable processes, common business rules, standard best practices, and custom methods and components tailored to your business will drive down the cost of implementing data integration projects. Aligning your world wide team around the same best practices pays large dividends over time.

The IBM Information Integration Solutions Group offers a collaborative program that enables customers to quickly build a Center of Excellence for data integration. Starting with a set of identified data integration projects and the IBM WebSphere Data Integration Suite, our Advanced Consulting Group uses a combination of structured workshops, onsite mentoring, and customized training to accelerate the completion of the tactical projects and results in a repeatable model for subsequent implementation projects.

In addition we offer coordinated training, mentoring, and certification for your System Integrators (SI) and development subcontractors. This provides you with an integrated project team that is certified in IBM best practices for enterprise data integration and ensures that they have the skills to leverage the IBM WebSphere Data Integration Suite.

The IBM COE offering provides guidance for quickly initiating the tactical projects; incorporates formal training at the appropriate time to maximize understanding and use of the products; and provides hands-on mentoring and assistance in the tactical development efforts. This approach delivers a working model of best practices and reusable methods with trained resources.

As this working model develops, the Information Integration Solutions Group's Advanced Consulting team will hold a series of workshops to develop:

- *Detailed project roles and responsibilities*
- *Best practice handbooks,*
- *Implementation guides*
- *Technical architecture recommendations*
- *Customized product training*

Together these deliverables form the infrastructure for a Center of Excellence.

Once the core components to the COE are developed, the Advanced Consulting Group will deliver deployment plans for the COE that detail:

- *The steps to deploying the COE in either a centralized or decentralized model*
- *A customized skills assessment tool and a certification plan for new staff, System Integrators, and development partners (onshore and offshore)*
- *Access to the IBM COE Portal that includes updates to best practices, a virtual consulting factory, and e-Learning courses*

IBM is committed to providing ongoing support to the COE team. As part of the Center of Excellence offering, customers may also take advantage of a Quarterly and Annual Review of the COE that includes a review of the best practices, technical architecture and training requirements.

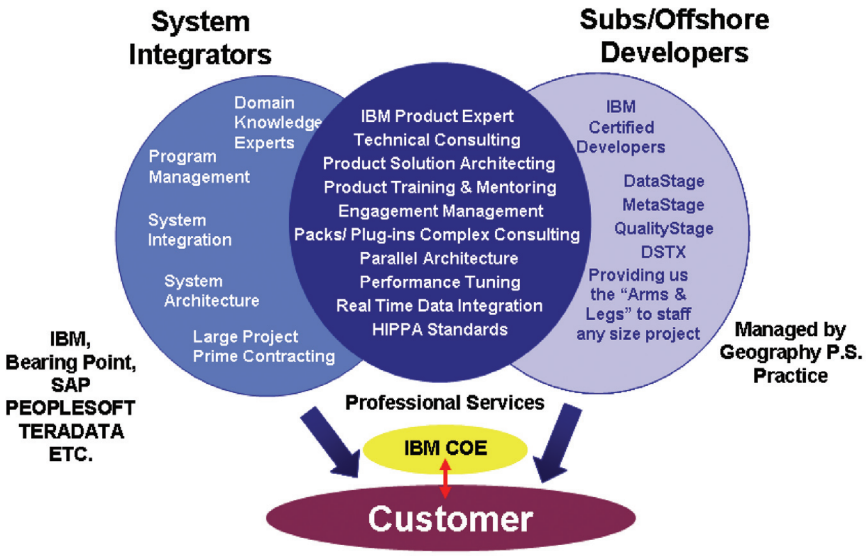

Figure 1. IBM's Center of Excellence for data integration is designed to incorporate system integrators and development partners into COE planning.

IBM can also provide staff augmentation as needed to fill the gaps in your technical organization, and help fill those gaps permanently through skill transition, training, and new hire mentoring. The COE offering from IBM's Information Integration Solutions Group accelerates the creation of a core competency in data integration within your organization. By leveraging the best practices and using the integration tools from IBM, you will drive down the costs of integration projects and deliver information that is accurate and more valuable to business users.

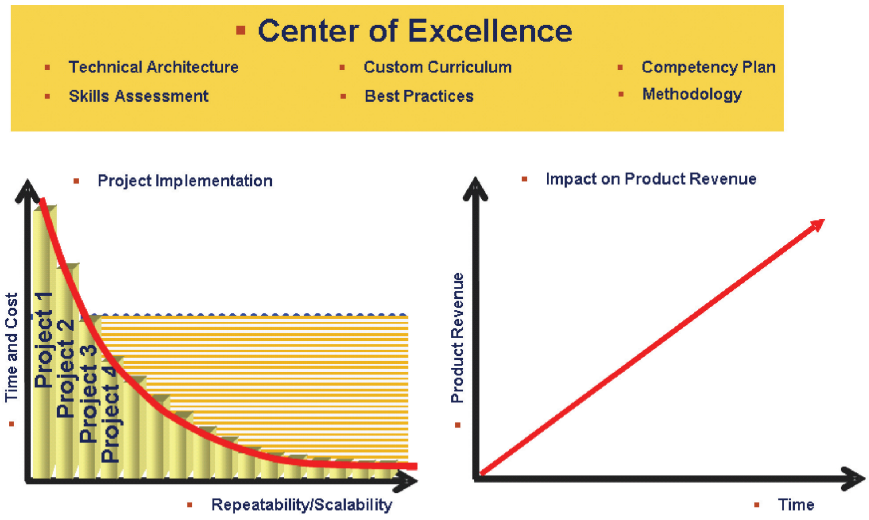

Figure 2. IBM's Center of Excellence for data integration enables companies to drive down the costs of information integration and bring new services and products to market faster.

Deploying IBM's Center of Excellence for data integration

The IBM Center of Excellence for data integration offering delivers customized:

- *Business plans for the COE that include timeframes and costs*
- *Workshops and on site mentoring for:*
  - *Technical architecture*
  - *Project management*
  - *Product usage*
- *Documented best practices and standards*
- *Robust, tool specific methodology*
- *Education and content re-licensing*
- *Global product support*
- *Staffing augmentation*
- *Access to an online community of experts and a knowledge base*
- *Access to virtual consulting and development resources*

The COE offering is designed to be delivered in a flexible way that meets your company's needs. The process of creating the COE is built around the delivery of identified data integration projects. The workshops and training are timed to meet your needs. This unique approach ensures that the COE team is applying the tools and methodologies in a mentored environment where a deep understanding of their business is combined with IBM's domain expertise and methodology for data integration to produce a world class competency for your company. Figure 3 presents a typical COE engagement process.

COE Deployment Model

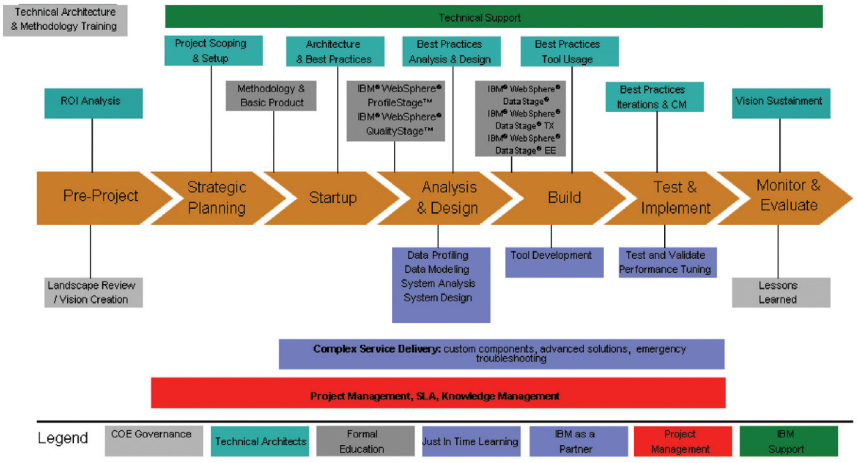

Figure 3. This diagram depicts a typical project progression with the various knowledge facets required to support the COE.

Center of Excellence workshops

The Center Of Excellence for data integration program from IBM is delivered in conjunction with a set of funded data integration projects. Throughout the projects, members of IBM's Advance Consulting Group will provide a series of workshops designed to uncover the requirements and best practices for establishing a COE within your IT organization.

First step: COE information exchange and discovery workshop

The first step in the program is for an IBM Solutions Architect and an IBM Technical Architect to conduct a two day information exchange and discovery workshop with the core data integration team. The key topics of this workshop are:

Create a Strategic Understanding of COE for data integration

- *Business awareness and support*
- *Review of all tactical and strategic projects*
- *Review of architectures and technologies*
- *Technical vision*
- *Organization approach*
- *Mission statement and shared vision*

Review Key Elements of the Center of Excellence

- *Defined roles and responsibilities*
- *Customized training requirements*
- *Skills assessment*
- *Discussion of Master Data issues*
- *Infrastructure management and progression*
- *Internal solution architect role*
- *Fostering reusability and custodianship*
- *Enterprise information models*

The specific deliverables from the Discovery Workshop is a complete proposal from IBM's Advanced Consulting Group including:

- *Descriptions of materials and deliverables for each phase of the project*
- *Plan and timings to create the COE*
- *Definition of roles*
- *Definition of IBM roles*
- *Investment plan*

Second step: setup, planning and requirements workshop

The second workshop is one week in duration onsite and develops a more detailed skills assessment and the full training requirements. The key topics in this workshop are:

Readiness Assessment, Skills Transition and Competency Planning

- *Agreement of COE management, structure, and operations*
- *Tactical and strategic just in time training and mentoring plan*
- *Full time staff certification plan*
- *Certification plan for key system integrators and offshore development partners*
- *Organization approach*
- *Customization planning for courses and best practices*

The specific deliverables from the Setup Workshop include:

- *COE operations Guide and Plan (customized with readiness assessment)*
- *Access to the COE Portal (10 seats)*
- *Access to e-Learning (10 seats)*

Third step: technical architecture and content workshop

The third workshop is two weeks in duration onsite and develops the architecture and deployment plan and instruction on methodology and best practices. The workshop is delivered by two Technical Architects and the key topics in this workshop are:

Architecture and Deployment

- *Technical review of architecture*
- *SOA, Web Services, and Real Time Messaging infrastructure review*
- *Computing platform, DASD storage, and Network recommendations*
- *Capacity and performance planning and discussion of Grid and HA*
- *Data integration product deployment review and guidance*

Methodology and Best Practices

- *Iterations II methodology, best practices, and curriculum review*
- *Repeatable processes, components, and rules processing review*
- *Curriculum customization to incorporate best practices*
- *Data quality and analysis review*
- *COE Portal, e-Learning, and Virtual Classroom training*

During this workshop the Technical Architects will work with your technical team (including the DBA and System Administrator) to install and, more importantly, configure all of the products to establish end-to-end connectivity.

- *Install and configure all products*
- *Validate end-to-end connectivity*
- *Ensure interoperability of all hardware and software components identified in the COE technical architecture*

The environment setup and testing begins the knowledge transfer process to the COE staff on product configuration and management. This mentoring process includes:

- *Reviewing all required software that is already installed*
- *Modification of product configurations for performance and scalability*
- *Estimating sizing requirements for all extraction stages*
- *Ensure connectivity to all source and target systems*
- *Assistance in resolving any conflicts with current installation and configuration*

Optional Elements

Optional: software sizing and capacity planning workshop

Customers may also choose to add a more in depth workshop focused on software sizing and capacity. This workshop and development of deliverables are estimated to require three weeks using two IBM Technical Architects.

The topics in this workshop are:

Sizing and Capacity Business Requirements

- *Review current technical infrastructure and architecture*
- *Develop simulation of customer environment*
- *Establish current capacity baseline*
- *Establish future capacity needs*
- *Develop recommended sizing and growth plan*

The specific deliverables from the Sizing and Capacity Workshop include:

- *Capacity planning tool-kit of scripts and executables for facilitating planning*
- *Customized sizing and capacity simulation model*
- *GAP analysis of current capacity requirements*

Optional: high availability planning workshop
Customers may also choose to focus on their requirements for a high availability implementation. This workshop and development of deliverables are estimated to require three weeks using two IBM Technical Architects. The topics in this workshop are:

High Availability Business Requirements

- *Review current technical infrastructure for applications needing high availability*
- *Review existing fail-over standards and software*
- *Review hardware configuration*
- *Map IBM product solutions to requirements*
- *Develop best practices for IBM products*

The specific deliverables from the High Availability Workshop include:

- *GAP analysis of current environment*
- *High Availability tool-kit of scripts and executables for HA deployments*
- *Customized HA best practices handbook*

Optional: Grid computing planning workshop
Customers may also choose to add a workshop specifically on a Grid implementation. This workshop and development of deliverables are estimated to require three weeks using two IBM Technical Architects. The topics in this workshop are:

Grid Computing Business Requirements

- *Review current technical infrastructure for proposed Grid applications*
- *Identify Grid software stack and resource management standards*
- *Review hardware configuration*
- *Map IBM product solutions to Grid requirements*
- *Develop best practices for IBM products*

The specific deliverables from the Grid Computing Workshop include:

- *GAP analysis of current environment for Grid computing*
- *Grid tool-kit of scripts and executables for Grid deployments*
- *Customized Grid implementation handbook for the target environment*

Center of Excellence for data integration handbooks, implementation guides, plans, and training
At the completion of the Center of Excellence for data integration program your organization will have completed a strategic project for the company, established an organization that is capable of configuring and managing your data integration environment, and have documented best practices, standards, and processes that can be deployed across the organization and shared with development partners.

COE handbooks:

- *Configuration and Performance Tuning*
- *Dataflow Job Design and Development*
- *Profiling and Data Quality*
- *Metadata Management*
- *Production Automation and Deployment*
- *Administration and Management*
- *Migration, Up-Grade, and Change Management*

These handbooks are a lasting resource of the COE that provide a common understanding on how the organization goes about the business of integrating enterprise data. Using and sharing the handbooks will help the COE staff leverage best practices, encourage reuse, and increase project team productivity.

In addition the COE staff will have completed a set of customized training classes with examples from your business and based on your data. The training classes provide hands-on experience with the entire IBM Data Integration Suite of products and are delivered during the project life cycle so that developers can apply the tools exactly when they first need them.

COE product training:

- *WebSphere DataStage® Enterprise Edition for extraction and transformation*
- *WebSphere ProfileStage™ for understanding data sources*
- *WebSphere QualityStage™ for data matching, cleansing, and standardization*
- *WebSphere MetaStage® for impact analysis and data lineage*

Along with the best practice handbooks and just in time training courses, the program delivers a set of plans, guides, and reusable components that are customized to your organization.

COE plans, guides and components:

- *The COE proposal and deployment plan*
- *Operations Guide and Plan*
- *Readiness Assessment and Skill Transition plan*
- *Enterprise Technical Architecture Recommendations*
- *Optional industry data models*
- *Optional custom components developed by IBM staff*
- *Certification program*

The IBM Center for Excellence for data integration program is also designed to provide your COE team ongoing access to information on the Iterations methodology, evolving best practices, and reusable data integration components.

COE portal:

- *Documentation for Iterations methodology*
- *All standard best practices documents*
- *DataStage Enterprise Edition jobs, maps, and data flow templates*
- *Custom components*
- *Custom brokers to extend MetaStage support for 3rd party product metadata*
- *Access to e-Learning*
- *Access to Virtual Consulting*

The COE always needs to operate at a high level of productivity. IBM offers quarterly and annual reviews with a Technical Architect to ensure that the project support continues as business requirements change.

COE quarterly checkup and annual review:

- *An assigned COE Account Manager*
- *Annual technical architecture review*
- *Quarterly development and production audits*
- *Ongoing access to e-learning (10 seats)*
- *Ongoing access to Virtual Classroom (10 seats)*
- *Standard Virtual Consulting Package*
- *Annual update of courseware and best practices to new releases of the IBM Data Integration Suite*

The IBM WebSphere Data Integration Suite

Companies also are looking for a standard technology platform for developing and deploying data integration processes. By adopting a standard set of tools, companies hope to:

- *Increase developer productivity*
- *Reduce the number of hand coded solutions that need to maintained*
- *Encourage the reuse of business rules and data transformations*

The IBM WebSphere Data Integration Suite supports all phases of an integration project life cycle including data profiling, data quality and data transformation. The WebSphere Data Integration Suite serves as the technology platform for the Center of Excellence.

Data profiling: The data profiling tools enable IT to understand what data is actually available to them by analyzing the content and the structure of each data source. Relying on outdated system documentation or simplistic data sampling to understand each source adds significant risks to integration projects. Setting realistic expectations for the business users about what the available data can actually do for them is critical at the start of any project. Without effective data profiling, you may find that you do not have the right data to meet their real business needs. You may also find that the cost of fixing data during the extraction and transformation process is very complex and adds too much time to the process. The result is that it will be difficult to meet the project deadlines and deliver the right data to the end users.

Data quality: The WebSphere Data Integration Suite also offers a complete solution for data quality. Often the most difficult part of a data transformation processes is the step where data is gathered from many disparate systems for the same business entity. If you need some data from multiple systems about the same customer, you need to be able to identify and match the customer in each system with customer data from the other systems, determine what data from the combined customer records should survive and what the standard format for that remaining data should be. The rules for matching and survivorship need to be jointly developed by the business and IT groups. This is an iterative process that if done well, can improve the quality of information dramatically. If done poorly, the new system will often become a point of contention because the users do not trust the data.

Data transformation: Of course, a new application requires that data from the source systems is transformed into new formats with new meaning that is appropriate for the new system. Having capabilities to connect, extract, transform, and deliver the data within the time frames available is critical to success. While some applications need data that is constantly being updated in small amounts, other applications require that large amounts of information be batch loaded on a scheduled basis. The processing needs to be scalable with demand from a single processor to an SMP environment to a full Grid environment without a requirement for changing code. That is why the WebSphere Data Integration Suite has the most comprehensive capabilities for connectivity, transformation, synchronization and parallel processing available today.

Metadata management: In each step of the data integration lifecycle, appropriate sharing and tracking of metadata is critical to maintain data lineage. The ability to track changes to data as it flows through the organization is become a central part of regulatory requirements. Equally important, tracking the metadata for semantic, technical, and project information is the key to ensuring that appropriate business and transformation rules are being applied. This information is critical for developers to perform an impact analysis of proposed changes across the data flow. By using the metadata that is shared across the WebSphere Data Integration Suite, it is possible to easily create data lineage reports and perform impact analysis to further improve the quality and consistency of information.

Summary

IBM's Integration Solutions Group has brought together everything required to create a fully functioning Center of Excellence for data integration. Our unique approach of developing the COE as we work with you on data integration projects ensures your success. The program is flexible, built and delivered according to the needs of your business. The program ensures that the right IBM experts- Solutions Architects, Technical Architects, Product Training and Services Managers, are onsite and available for mentoring when you need them. The deliverables include completed data integration projects as well as the plans, handbooks, guides, and methodology that form the infrastructure for the core competency of your COE staff. The Center of Excellence Offering gives you a practical approach to mastering enterprise data integration with measurable ROI and a long term commitment to your success.

Next steps

To learn more about the Center of Excellence, please contact your IBM marketing representative or visit **ibm.com**/software/data/integration

© Copyright IBM Corporation 2005

IBM Software Group
Route 100
Somers, NY 10589
U.S.A.

Produced in the United States of America
11-05
All Rights Reserved

IBM, the IBM logo, the On Demand Business logo, AuditStage, DataStage, MetaStage, ProfileStage, QualityStage and WebSphere are trademarks of International Business Machines Corporation in the United States, other countries or both.

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 IBM home page on the Internet can be found at **ibm.com**

GC-18-9748-00