Delivering information you can trust

June 2008

Information Management software

# IBM InfoSphere Information Server: Simplify integration with unified metadata

Contents



## The information integration and collaboration challenge

Information integration is an extraordinarily complex activity that affects every part of an organization. It is the deciding factor behind a business' success or failure—and in many cases its very survival. Today, organizations face a wide range of information-related challenges: varied and often unknown data quality problems, disputes over the meaning and context of information, managing multiple complex transformations, leveraging existing integration processes rather than duplicating effort, ever-increasing quantities of data, shrinking processing windows and the growing need for monitoring and security to ensure compliance with national and international law.

To further complicate the situation, many businesses have implemented a seemingly endless set of disparate tools to combat the integration problem. Deploying these tools together as part of a single mature, manageable process is critical to the success of the business. However, organizations often spend more time integrating the technologies they purchase and defining new processes to use these fragmented solutions than integrating their own data to solve real business problems.

The manual nature of the integration process makes not only development but also collaboration inefficient. In a typical integration workflow, a data modeler begins by defining a target data model. The data modeler's structures are then sent to business and data analysts, who manually define source-to-target mapping specifications based on their understanding of the data profile—which

is sometimes correct and sometimes incorrect. The mapping specification is often stored in a Microsoft® Excel® spreadsheet, which is printed and sent to the developer to be manually converted into an extract, transform and load (ETL) job. The lack of automation and audit trails in these labor-intensive processes slows down the project, increases the potential for errors and minimizes the possibility of component reuse by other staff members or in new projects.

### Streamlining development with IBM InfoSphere Information Server unified metadata

Leveraging its years of experience with thousands of customer implementations, IBM has removed these inherent obstacles and created a unified information integration platform built on a services-based architecture and a single, active metadata repository. This allows businesses to focus on integrating the data to drive business value, rather than on integrating multiple vendor products that were never designed to be used together.

The IBM® InfoSphere™ Information Server unified metadata platform enables companies to deliver on three key success factors critical for true information integration:

- *Collaboration*
- *Trust*
- *Compliance*

To address these factors, an enterprise must take a holistic approach to the traditional system-development life cycle. The enterprise must consider not just technology, but the people and associated processes as well. Organizations must be able to access and share the wealth of business, technical and operational metadata generated across disparate communities to complete the enterprise picture and deliver trusted information on demand.

The patented metadata repository and services of the IBM InfoSphere Information Server platform are designed to support a unified metadata strategy that helps streamline communication and collaboration and simplify the delivery of enterprise projects. IBM InfoSphere Information Server enables organizations to seamlessly store, enhance and exchange metadata, which is generated as a natural consequence of the data integration process, to automatically maintain consistency across projects and teams.

***Three approaches to information integration***

Many software vendors have implemented rudimentary product integration within their tool offerings to display information from one product interface within another in a read-only context. This approach falsely provides the perception that metadata is being actively shared across disparate tools—each with completely different back-end architectures and metadata storage mechanisms.

In reality, this superficial interface integration does not facilitate true collaboration-driven development or generate new insights to help maximize the end-to-end integration life cycle. A few software vendors have attempted a deeper level of integration called engine integration—the consolidation of multiple integration processes into a single processing engine designed to scale as projects grow in scope and data volume size. Until recently, however, no vendor had successfully attempted metadata repository integration to support a true enterprise infrastructure platform offering.

Metadata repository integration is significantly more complex and difficult because it requires the organization to gather information spread across multiple processes, consolidate it into a single storage area and then reconnect the information with multiple user viewpoints to deliver additional value to downstream users and processes. While a difficult problem to solve, being able to instantly share information across not only multiple user roles and tasks but also across multiple integration processes and projects allows companies to take full advantage of their investments and streamline the overall efficiency of their development process. The IBM InfoSphere Information Server platform is designed to facilitate this type of meaningful integration, allowing organizations to focus on solving complex business problems rather than integrating separate profiling, cleansing and ETL data technologies.

IBM InfoSphere Information Server features a unified set of product modules designed to streamline the process of building a data integration application (see Figure 1). The IBM InfoSphere Information Server platform offers a comprehensive, integrated architecture built upon a single shared metadata repository—allowing information to be shared seamlessly among project data integration tasks. Organizations can use information validation, access and business processing rules across multiple projects, leading to a higher degree of consistency, greater control over data and improved efficiencies.

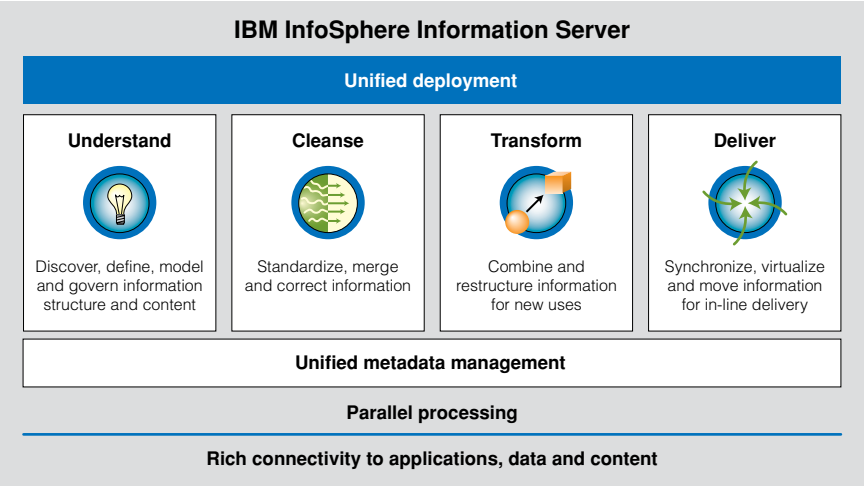



*Figure 1: IBM InfoSphere Information Server is a flexible data integration platform designed to deliver trusted information on demand for key business initiatives.*

IBM InfoSphere Information Server enables businesses to perform five key integration functions:

1. ***Understand the data.*** *IBM InfoSphere Information Server can help companies automatically discover, model, define and govern information content and structure, as well as understand and analyze the meaning, relationships and lineage of information. With these capabilities, organizations can better understand data sources and relationships and define the business rules that eliminate the risk of using or proliferating bad data.*
2. ***Cleanse the data.*** *IBM InfoSphere Information Server supports information quality and consistency by standardizing, validating, matching and merging data. The platform can help companies create a single, comprehensive, accurate view of information by matching records across or within data sources and allowing a single record to survive from the best information available.*
3. ***Transform data into information.*** *IBM InfoSphere Information Server transforms and enriches information to help ensure that it is in the proper context for new uses. It also provides high-volume, complex data transformation and movement functionality that can be used for stand-alone ETL scenarios or as a real-time data processing engine for applications or processes.*
4. ***Deliver the right information at the right time.*** *IBM InfoSphere Information Server provides the ability to virtualize, synchronize or move information to the people, processes or applications that need it. It also supports critical Service Oriented Architectures (SOAs) by allowing transformation rules to be deployed and reused as services across multiple enterprise applications.*

5. ***Perform unified metadata management.*** *IBM InfoSphere Information Server is built on a unified metadata infrastructure that enables shared understanding between the different user roles involved in a data integration project, including business, operational and technical domains. This common, managed infrastructure helps reduce development time and provides a persistent record that can improve confidence in information while helping to eliminate manual coordination efforts.*

### *Support multiple user roles through unified metadata*

Each user role requires access to certain metadata to perform assigned tasks. These same users must also be able to seamlessly and directly share information with other users in different roles, so streamlining collaboration among multiple, disparate users is critical to the success of any integration effort.

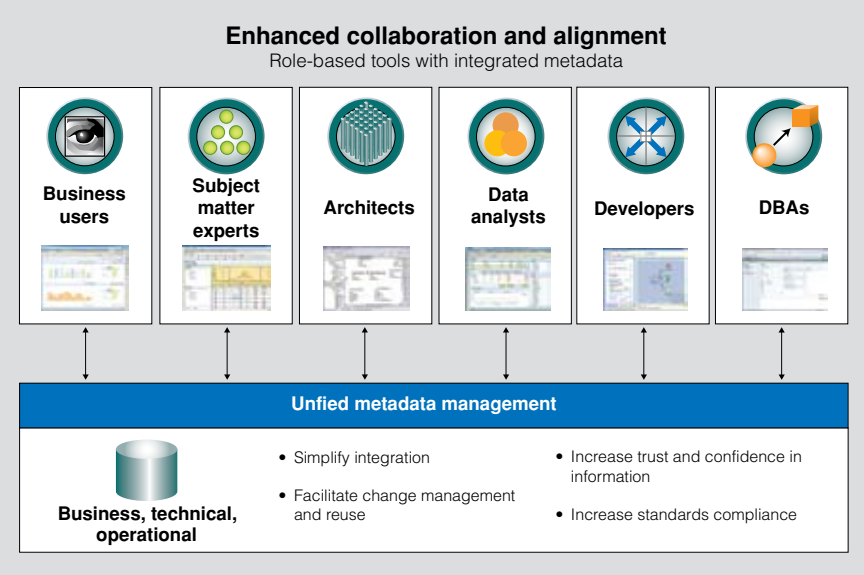
Metadata user roles typically fall into several categories:

- ***Project managers*** *have overall responsibility for the project, including training, deployment, management, resource allocation, coordination and progress tracking.*
- ***Administrators*** *oversee key areas of system implementation. Database administrators (DBAs) are in charge of database commissioning, installation, configuration, deployment and management, as well as data modeling. System and software administrators are responsible for project hardware, software, configuration, environment maintenance and deployment, as well as management of system users and their associated security roles.*

- ***Business analysts*** *provide in-depth business background and context to ensure that integration implementations meet business requirements and needs. Subject matter experts interpret end-user business requirements and assist the project team in prioritizing requests, defining business terms for context, documenting business transformation logic and developing the project validation criteria.*
- ***Data stewards*** *have deep business and technical knowledge that often bridges the gap between business and IT. Stewards manage logical data resources and coordinate data definitions, aliases, quality control, improvement efforts, access authorization and planning for subject area data.*
- ***Architects*** *help ensure that enterprise standards are met and applied consistently across the enterprise and across projects. Information architects often manage the process to standardize efforts related to metadata creation, maintenance, enhancement and distribution. They also work with developers to advocate data integration best practices in the development teams.*
- ***Integration developers*** *create processes and jobs to manage data manipulation. For example, data cleansing developers design and develop complex data cleansing applications. Data integration developers create and test ETL applications to support data integration and cleansing solutions, and SOA developers deploy data services and ensure that existing services are reusable and will meet service level agreements (SLAs). SOA developers also create service documentation for access by other members, as well as maintain services and registry information.*

The majority of these user roles are involved in every data integration project—sometimes with a single user performing multiple roles. Each user generates critical metadata as a natural consequence of the specific task being performed. These tasks are often performed in parallel. Because each IBM InfoSphere Information Server platform module is specific to a subset of users and their designated tasks, users work only with the metadata relevant to the job they are performing (see Figure 2).

*Figure 2: Metadata generated and consumed during the integration process is role- and task-based.*



***Industry-leading metadata architecture facilitates flexible, highly scalable integration***

IBM's history of developing innovative technology solutions extends back to familiar mainframe systems and enterprise application integration technologies, such as MQ Series. In the 1990s, IBM released its consolidated application server, IBM WebSphere®, and remains a market leader according to the Gartner Magic Quadrant for Application Infrastructure, Q207.[1] IBM InfoSphere Information Server, a flagship unified data integration platform, was launched in November 2006. IBM remains a leading visionary vendor for integration technology according to the Gartner Magic Quadrant for Data Integration Tools, 2007.[2] By architecting the core data integration components into a single platform through repository, engine and interface integration (see Figure 3), IBM created a comprehensive information integration platform while also protecting companies' prior technology investments.

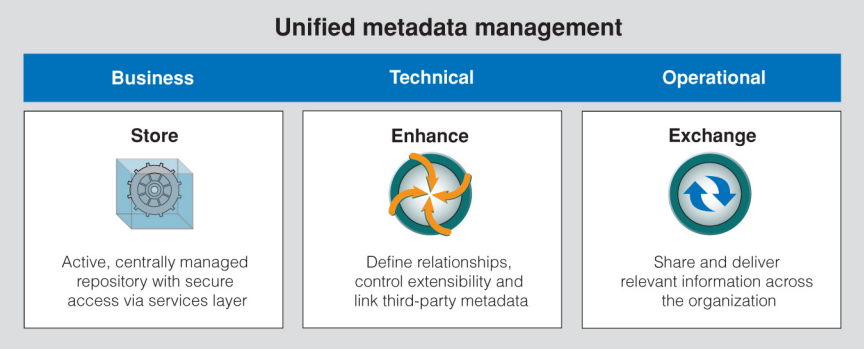



*Figure 3: The IBM InfoSphere Information Server unified metadata management architecture is a single, active metadata repository designed to support flexible, highly scalable integration requirements.*

IBM InfoSphere Information Server supports three primary types of metadata: business, technical and operational.

**Business metadata** is critical to providing context for an integration project. It helps define terms in everyday language, without regard for technical implementation. For example, the language used to describe what a customer is and how to categorize a customer is often business-specific and may differ between company divisions.

**Technical metadata** is often used by more technical staff, such as developers. It includes items such as table definitions and data types. These objects are used heavily during the application design and development process.

**Operational metadata** refers to the metadata generated and captured when a process executes. It allows administrators to manage the system and ensure things are running smoothly; it also helps them troubleshoot issues if there is a problem with a process.

Unifying these types of metadata creates an end-to-end relationship, enabling users to understand not just where information is stored and what happened to it as it moved through the organization, but also the business context of that information.

The IBM InfoSphere Information Server unified metadata repository provides the architecture to support three key information-related tasks: store, enhance and exchange.

1. ***Store:*** *The unified metadata foundation provides a single active repository to facilitate shared understanding across business and technical domains. By using a single shared active repository, metadata creation and maintenance becomes a natural consequence of using the integration platform components—thus removing the administrative burden of manual metadata consolidation and management. Active metadata sharing also enables developers to work more productively, helps to improve visibility and manageability of the organization's information assets and fosters collaboration.*

   *Metadata is accessed via an internal services layer to facilitate scheduling, security, logging and error handling, reporting, analysis and search functions. By consolidating these tasks into common services, IBM InfoSphere Information Server can offer outstanding out-of-the-box scalability, security and integration. Exploiting these services as part of an application server helps provide data and code integrity, centralized configuration, security, performance and a lower total cost of ownership (TCO).*

2. ***Enhance:*** *Metadata stored within the common repository can be enhanced through linkages to third-party metadata that is a direct part of the integration flow, such as business intelligence (BI) and data modeling tool metadata. This metadata can be imported into the repository, and linkages and relationships between it and IBM InfoSphere Information Server can be created to expand the understanding of metadata across organizational processes. In addition, users can control metadata extensibility to capture unique business metadata that requires tracking.*
3. ***Exchange:*** *A delivery and exchange mechanism is critical for exposing metadata to different user roles involved in integration projects. IBM InfoSphere Information Server includes task-driven modules for this purpose, as well as a Web-based, consolidated reporting layer that operates across modules and generates metadata reports to be accessed via a common Web-based console. Reports can be scheduled to run automatically and access can be controlled per user. PDF, HTML, RTF and text versions of reports can be produced for historical purposes.*

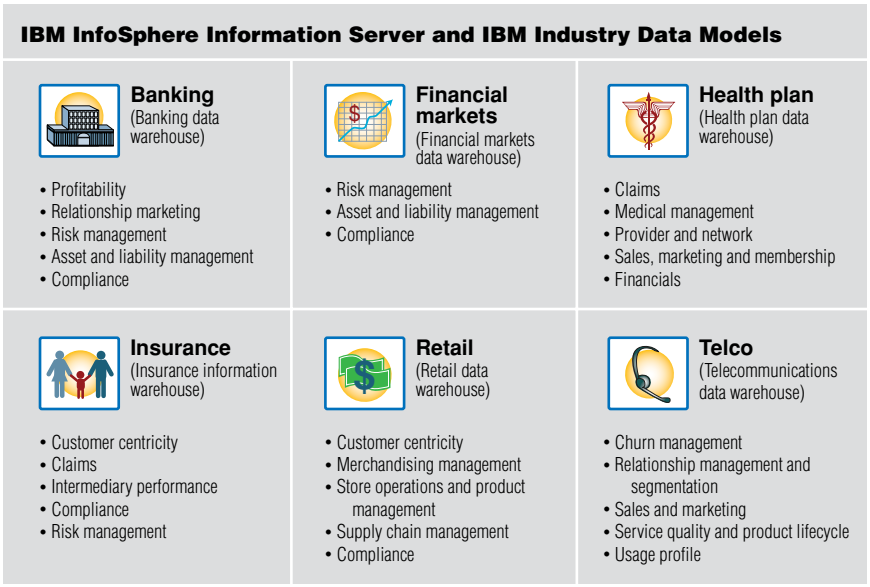
### IBM InfoSphere Information Server integrated modules

The IBM InfoSphere Information Server platform consists of multiple modules that can be deployed together or individually within an enterprise integration framework. The following sections describe the components and the metadata they generate, consume and share.

### *IBM Industry Data Models and IBM Rational Data Architect*

Both business and IT can benefit from using the IBM Industry Data Models to help implement key strategic business initiatives faster, more reliably and confidently. Developed using IBM's experience with more than 400 clients and more than 10 years of development expertise, the IBM Industry Data Models uniquely support the six major industry verticals (see Figure 4).

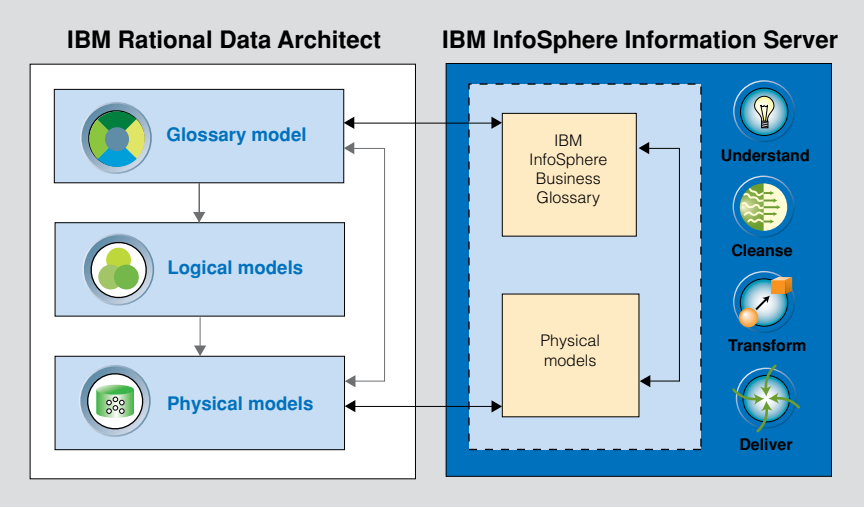
*Figure 4: IBM Industry Data Models, built based on years of experience, cover all major market verticals.*



The IBM Industry Data Models include a glossary of terms and concepts and a physical data model. This information, stored in a data modeling tool such as IBM Rational® Data Architect, can be shared with IBM InfoSphere Information Server to better align business and IT users and help accelerate project delivery.

IBM Rational Data Architect is just one of the major data modeling tools integrated with IBM InfoSphere Information Server. Rational Data Architect acts as a modeling gateway into the unified metadata management capabilities of IBM InfoSphere Information Server, interchanging glossary and physical metadata (see Figure 5). This metadata is then exposed and leveraged across each IBM InfoSphere Information Server module as necessary.

*Figure 5: Rational Data Architect allows companies to publish metadata such as business glossaries and physical data models to IBM InfoSphere Information Server.*



Rational Data Architect distinguishes between three types of models that allow organizations to develop concepts at different levels of abstraction. Users can easily switch between these three types using defined relationships to maintain consistency:

- *A **glossary model** describes the business terminology that is used within an organization and the hierarchies and relationships inherent in that terminology.*
- *A **logical data model** describes abstract entities about which an organization wants to collect data and the relationships among these entities.*
- *A **physical data model** is a database-specific model that represents relational data objects and their relationships. Companies can use a physical data model to generate data definition language (DDL) statements, which can then be deployed to a database server.*

The tight integration between the IBM Industry Data Models, Rational Data Architect and IBM InfoSphere Information Server allows organizations to exploit industry-specific business and technical metadata to accelerate data integration projects such as master data management initiatives or data warehouse development. For example, the Industry Data Models and Rational Data Architect physical schemas can be shared across the entire IBM InfoSphere Information Server platform, including InfoSphere Information Analyzer, InfoSphere FastTrack, InfoSphere DataStage® and InfoSphere QualityStage®. In addition, business or glossary definitions from the Industry Data Models and Rational Data Architect can be used to populate InfoSphere Business Glossary to share common definitions across the enterprise.

***IBM InfoSphere Business Glossary***

A business glossary, sometimes called a dictionary, contains definitions of terms used by the organization to support business initiatives. The glossary defines the language of the enterprise and, by extension, the language of projects—providing the collaboration link between disparate groups involved in integration efforts. Without a formal glossary to capture and centrally manage this valuable company asset, organizations run the risk of this critical information leaving the building each day when employees go home.

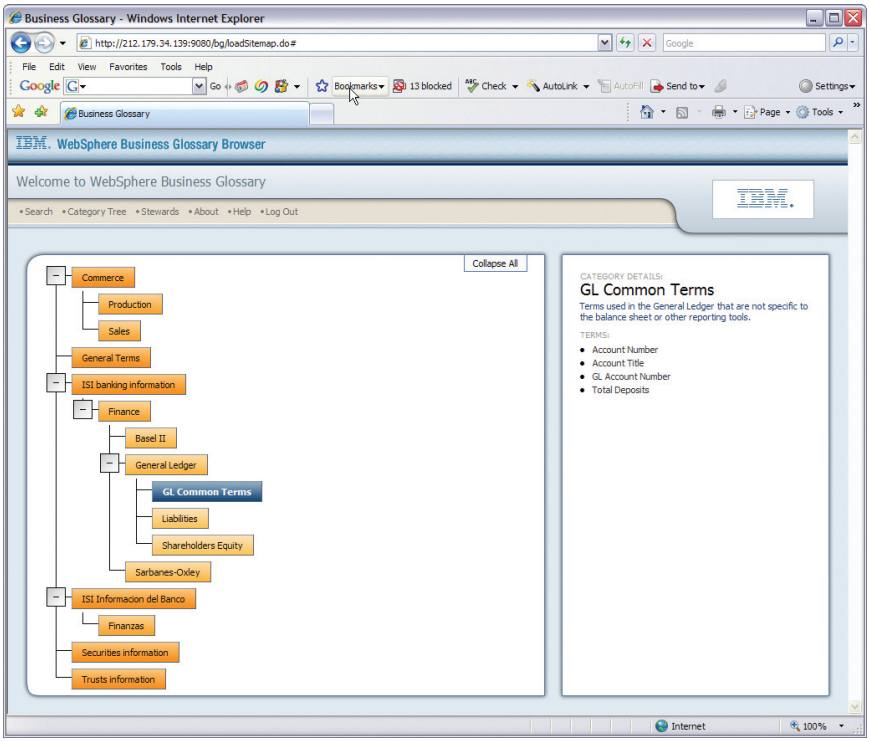
IBM InfoSphere Business Glossary enables data analysts, business analysts and subject matter experts to create a rich glossary, linking business concepts to technical metadata and exposing these linkages across the entire enterprise through easy-to-use, simple interfaces. InfoSphere Business Glossary facilitates the creation of comprehensive, authoritative terms and defined relationships via categorical hierarchies. Organizations can assign data stewards to manage this information to support data governance initiatives that require accountability and responsibility for topics, assets and relationships. InfoSphere Business Glossary users have direct access to the steward's contact information—including name, phone number and e-mail address—so they know the person to contact when they have a question or requirement concerning a business term.

InfoSphere Business Glossary includes three interfaces designed for specific user audiences:

1. ***Business Glossary:*** *Designed for the data steward, Business Glossary enables subject matter experts to create rich, detailed definitions for terms and define categories to represent the relationships between the terms. To describe organization-specific properties about particular terms of relevance, users can also add custom attributes that extend the meaning of items in the glossary.*

   *In addition, business analysts can use this interface to link technical artifacts such as database tables and columns to business terms. This linking helps ensure that particular data artifacts are coupled with their business context and enables two-way communication. Business users can drill down from a term to find the technical data sources, while technical users working on a data source or ETL job can understand the business context of the objects being used.*

*Figure 6: InfoSphere Business Glossary facilitates business and IT communications by creating and managing a common business vocabulary.*




2. ***Business Glossary browser:*** *Designed for the business user, Business Glossary browser is an intuitive, read-only browser interface that does not require training (see Figure 6). Business users can search and explore the vocabulary and its classification of data assets, identify stewards responsible for assets and provide direct feedback on business information.*
3. ***Business Glossary Anywhere:*** *Designed to allow anyone in the organization to view the contents of the common glossary and to promote the adoption of a standardized language across the enterprise, Business Glossary Anywhere is invoked directly from any application. Users can search any term without losing the context of the application they are currently using. A single click produces a small window with information about associated metadata in the business glossary, including the steward of the term.*

In addition to using InfoSphere Business Glossary to create and edit the glossary contents, companies can import metadata from other sources, such as IBM Rational Data Architect or .csv and XML files. These additional mechanisms help populate the business glossary and eliminate the manual input of metadata, which promotes consistency of business terms and helps reduce the chance of introducing errors. Organizations can also use the IBM Industry Data Models to seed InfoSphere Business Glossary with industry-standard terms.

The collaborative management of InfoSphere Business Glossary—coupled with the glossary capabilities of Rational Data Architect and rich business content of the IBM Industry Data Models—provides a comprehensive solution for building an enterprise glossary. But just as important as building and managing that glossary is the ability to expose the content to different users across an organization. An understanding of the business context allows organizations to create better applications to meet business requirements. As such, InfoSphere Business Glossary content is exposed to other IBM InfoSphere Information Server modules including InfoSphere Information Analyzer and InfoSphere FastTrack. Users of these applications possess the organizational knowledge and capability to help bridge the gap between business and technical metadata and to establish relationships defining these linkages. InfoSphere Business Glossary terms are directly shared with each of these modules so users can create additional relationships and create extended content for the common glossary as part of their data integration tasks and workflow processes.

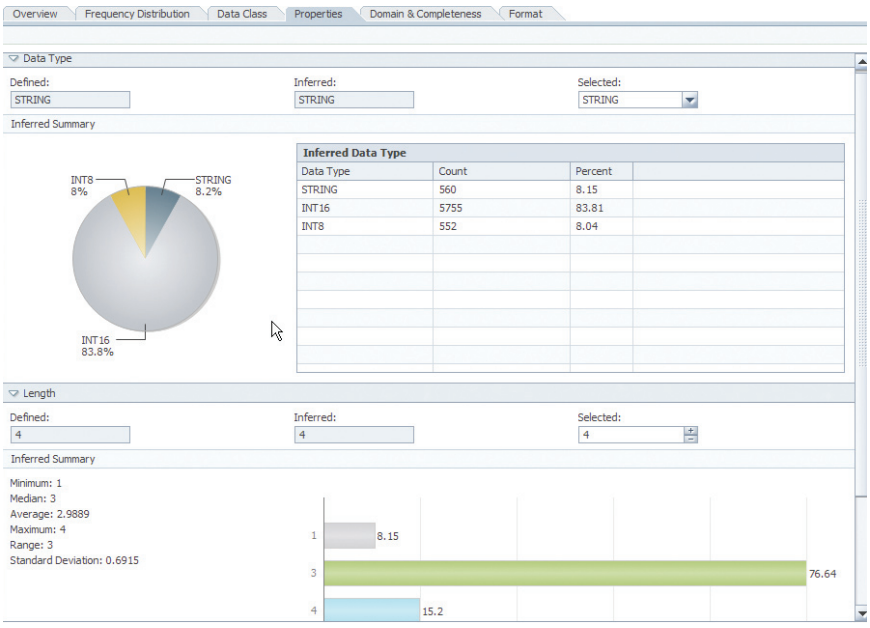
***IBM InfoSphere Information Analyzer***

Over time, data stored in legacy systems and enterprise applications can lose much of its value as metadata, field usage and general knowledge about data changes. Although the data may have served its original purpose perfectly, it is often inappropriate for use in other applications as time passes and requirements evolve.

The IBM InfoSphere Information Analyzer module provides data profiling capabilities that can help reduce project costs and risk by discovering problems in the early stages and monitoring changes in data structure and content. Data profiling techniques focus on the underlying content of the data: examining the values in each column, inferring additional information from those values and then leveraging that information to assess structural integrity and relationships across data sources. InfoSphere Information Analyzer provides a comprehensive view into a company's data assets through a repeatable, proven process.

InfoSphere Information Analyzer starts by capturing existing definitions of schemas, tables, files and columns and making that information available to the rest of the IBM InfoSphere Information Server platform (see Figure 7). Column analysis looks explicitly at distinct values, allowing the tool to generate inferred metadata and determine the true physical characteristics of the data. Furthermore, primary key, foreign key and cross-domain analysis processes help highlight duplicated data, broken keys or missing or invalid data relationships across tables—problems that can affect business processing, data migration efforts or the loading of critical data into production systems or data warehouses. Analysts working with these results can enrich the metadata by creating annotations that can be shared with IBM InfoSphere Information Server modules; evaluating differences between defined and inferred

*Figure 7: IBM InfoSphere Information Analyzer can perform column analysis to generate inferred metadata and determine the true physical characteristics of the data.*



metadata; identifying incomplete or invalid values; generating reference and mapping tables for use in IBM InfoSphere DataStage or IBM InfoSphere QualityStage; and linking the physical data to the semantic terms entered through InfoSphere Business Glossary.

InfoSphere Information Analyzer fulfills a critical role in the integration process. Profiling data helps the analyst to completely understand the data sources prior to starting the design of the detailed mapping specifications. To facilitate the most complete and accurate mapping specifications, profiling results from InfoSphere Information Analyzer are directly accessible from within InfoSphere FastTrack, where the specifications are defined and documented. These specifications then become the input requirements for the InfoSphere DataStage and InfoSphere QualityStage ETL and cleansing jobs that support the business application being developed. The more information the analyst has about the true data structures and content, the more accurate the requirements are for the downstream developers.

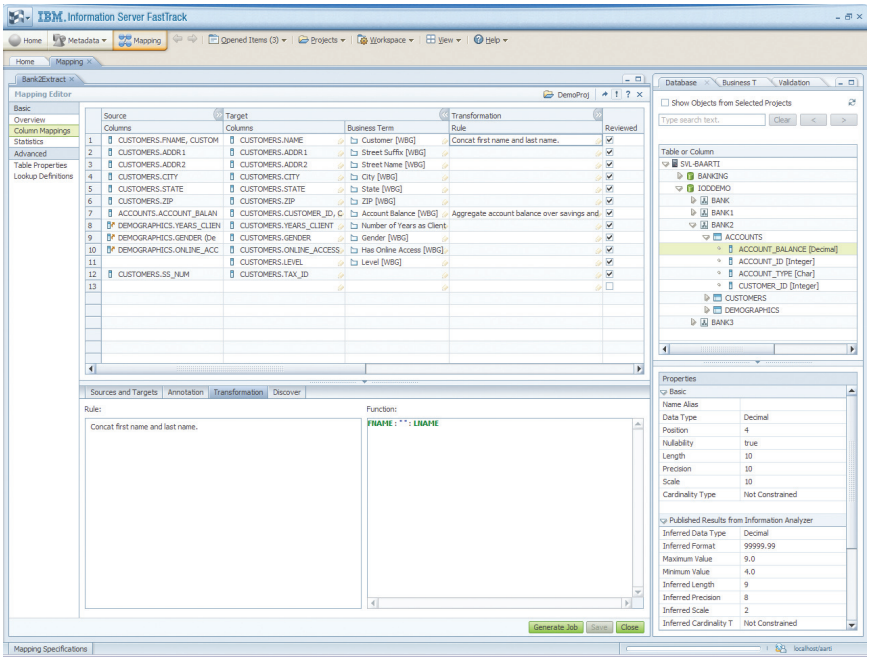
***IBM InfoSphere FastTrack***

During integration projects, companies often spend an inordinate amount of development time overcoming differences in languages, skills and working methods; clarifying business requirements; and synchronizing tool output. IBM InfoSphere FastTrack is designed to break down these barriers to help maximize team collaboration, increase automation and ensure that projects are completed successfully and on time.

By creating an integrated environment that includes business analysts, developers and data modelers, IBM InfoSphere FastTrack accelerates collaborative development across user roles, products and geographies. It is designed to automate efforts across multiple data integration tasks while incorporating business perspective and maintaining lineage and documented requirements.

IBM InfoSphere FastTrack captures and stores critical metadata about key business requirements to streamline the application development process. Using InfoSphere Information Analyzer, data analysts and subject matter experts profile data from multiple sources to understand the types of transformation rules that need to be applied during the migration process.

*Figure 8: IBM InfoSphere FastTrack creates an integrated environment that includes business analysts, developers and data modelers.*




The metadata created during profiling is directly accessible within IBM InfoSphere FastTrack and includes information about inferred data types, data lengths, data values and primary and foreign keys—as well as any specific notes entered by the data analyst (see Figure 8). Using this profiling information, analysts create source-to-target mapping specifications that describe how to extract, combine and transform information to meet business requirements. The source and target metadata can be accessed directly or may be imported from the IBM Industry Data Models or Rational Data Architect.

IBM InfoSphere FastTrack also enables companies to incorporate a business semantic layer to help analysts create and define relationships between the business terminology and the physical representation layer. Analysts can leverage the contents of InfoSphere Business Glossary or create and publish new business terms to be included in the project documentation. Once the mapping specification is complete, any updates to the target model structure are also included in the specification details. The analyst can then generate InfoSphere DataStage and InfoSphere QualityStage jobs and hand them off to the developer to review and complete for production deployment.

***IBM InfoSphere DataStage and InfoSphere QualityStage***

The InfoSphere DataStage module is designed to provide the functionality, flexibility and scalability required to perform data transformation functions for complex data integration initiatives. With the ability to manage multiple integration processes, InfoSphere DataStage enables direct connectivity to enterprise applications as either sources or targets of metadata. Transformations can be processed in batch, in real time, in near real time or as part of a service oriented architecture.

InfoSphere DataStage leverages the parallel execution capabilities of IBM InfoSphere Information Server to meet companies' most demanding data volume and transformation requirements. It has been certified by a third-party benchmark auditor accredited by the Transaction Processing Performance Council to scale almost near linearly as hardware is added to the processing environment.[3] In addition to running on UNIX®, Microsoft Windows® and Linux® platforms, InfoSphere DataStage also runs natively on mainframe IBM S/390® and IBM System z™ Linux environments, enabling organizations to fully leverage their IT investments.

InfoSphere DataStage ETL jobs comprise technical design metadata that describes the job flow and transformation logic being applied. Operational or run-time and parameter metadata is also captured when the jobs are executed. IBM InfoSphere Information Server has the unique capability of linking the design and operational metadata together to provide a complete picture of what actually happens when a job executes in a production environment at 2 A.M. This is critical for supporting and troubleshooting complex integration environments, as well as supporting compliance reporting for tracking data lineage.

IBM InfoSphere QualityStage complements InfoSphere DataStage, enabling developers to create in-line data cleansing processes as components directly within the DataStage ETL canvas (see Figure 9). Using the stages and design

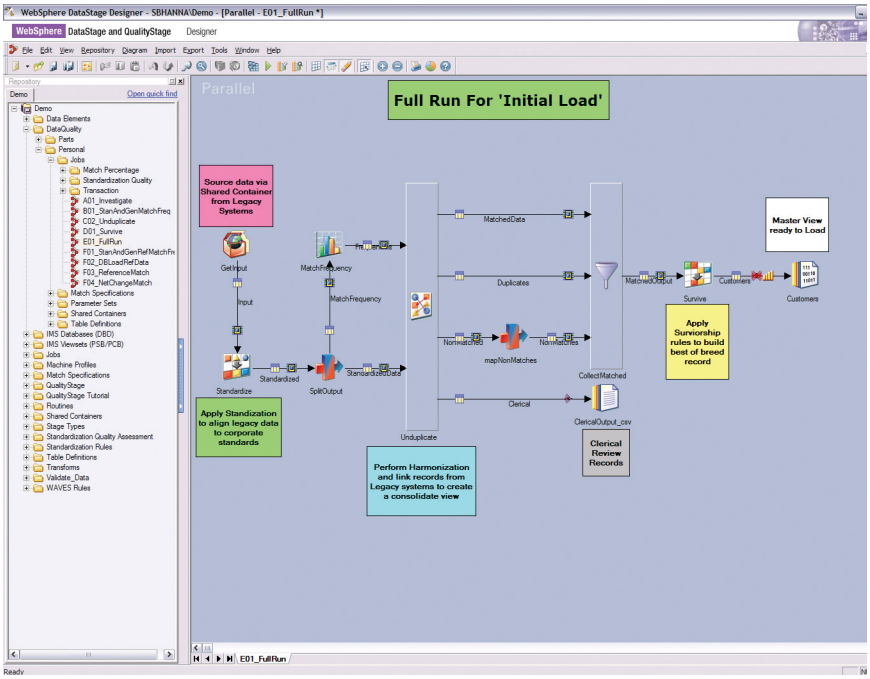



*Figure 9: InfoSphere DataStage and InfoSphere QualityStage enable developers to create in-line data cleansing processes.*

components, developers can quickly and easily process large stores of data while cleansing and combining sources as needed. The probabilistic data matching capabilities and dynamic weighting strategies within InfoSphere QualityStage support the creation of high-quality, accurate data by consistently linking and consolidating core business information—such as customer, location and product—throughout the enterprise. This helps reduce the time and cost of implementing strategic, domain-focused projects by improving organizations' understanding of their master data. The business rules defined to combine and consolidate records are stored in the metadata repository and shared with other modules of IBM InfoSphere Information Server for full insight and auditing into the cleansing and data consolidation processes.

By storing the ETL, cleansing, design and operational metadata in the InfoSphere Information Server common metadata repository, organizations can follow the linage of information through the entire integration process. This understanding facilitates trust in the end-user application and helps meet regulatory compliance requirements for traceability and proof of pedigree.

Because it is not always feasible to move data to a centralized location, developers can use IBM InfoSphere Federation Server to perform queries across disparate data sources. With this module, organizations can virtualize their data and provide information in a form that applications and users can leverage, yet hides the complexity of the underlying sources. These federated queries can be exposed directly within an ETL job flow to help expand connectivity and simplify the transformation design flow.

In addition to traditional batch processing and federated query support, IBM InfoSphere Information Server also supports change data capture (CDC) and replication. Organizations often need to process data immediately as soon as it changes without delay. InfoSphere DataStage can source information from InfoSphere Change Data Capture using log-based CDC technology to provide scalable, high-performance and heterogeneous data integration without impacting the source systems. The option to leverage real-time data integration allows customers to get the information they need when they need it to help them to make decisions at the speed of business.

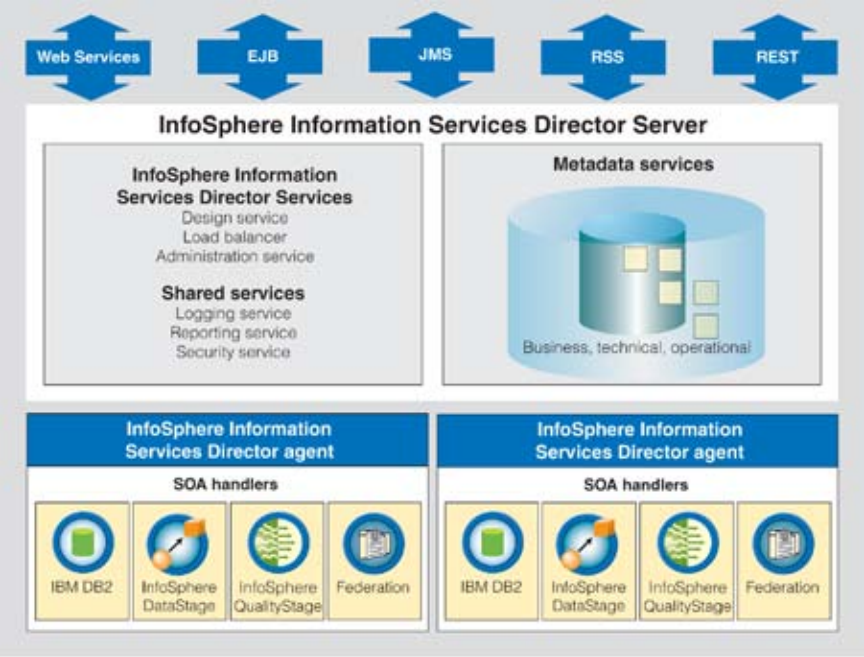
InfoSphere Change Data Capture for Replication products are designed for homogeneous environments that require a powerful solution for data distribution, data integration and data availability to provide a consistent view of information between systems. High performance and scalable replication ensures that primary and secondary systems are synchronized in real time.

***IBM InfoSphere Information Services Director***

Publication of consistent, reusable services makes it easier for business processes to get the information they need from across a heterogeneous IT landscape. Using InfoSphere Information Services Director, developers can expose InfoSphere DataStage, InfoSphere QualityStage, InfoSphere Federation Server, InfoSphere Classic Federation Server for z/OS® and IBM DB2® logic as services that are deployed and shared across the enterprise for application and process integration. IBM InfoSphere Information Services Director load balances these service requests across multiple IBM InfoSphere Information Server nodes to help ensure fault tolerance and high availability.

The InfoSphere Information Services Director tool packages information integration logic as services that are designed to insulate developers from underlying sources, and it allows these services to be invoked as Enterprise JavaBeans™, Java™ Message Service (JMS), Web services, Really Simple Syndication (RSS) and Representational State Transfer (REST) (see Figure 10).

*Figure 10: InfoSphere Information Services Director helps balance service requests across multiple IBM InfoSphere Information Server nodes.*



Information about the design of InfoSphere Information Services Director applications and the run-time deployment of applications is stored within the common metadata repository of IBM InfoSphere Information Server. This metadata enables a variety of important services:

- ***Infrastructure services*** *including logging, security, information service cataloging and load balancing and availability services*
- ***Information provider handling*** *for providers including IBM DB2, InfoSphere Federation Server, InfoSphere Classic Federation Server for z/OS, InfoSphere DataStage, InfoSphere QualityStage, InfoSphere Master Data Management (MDM) Server and Oracle*
- ***Service bindings*** *that allow service consumers to access information services using multiple technologies for program interoperability (bindings)*

SOA environments focus on decoupling the service from the underlying process so users do not have to understand the complexities of the process being performed. With IBM InfoSphere Information Server, Web developers have the option to create services as well as see "under the covers" for troubleshooting because the metadata is maintained in one common area. The created services can be exposed within the IBM InfoSphere Information Server Web console for common access as well as in other IBM products, such as InfoSphere Service Registry Repository.

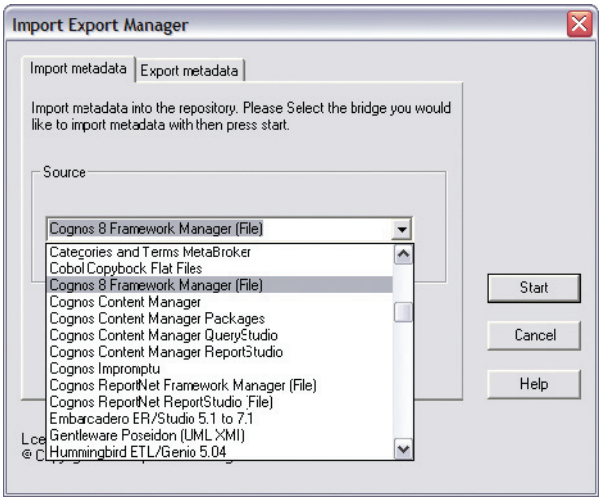
#### *IBM Import Export Manager*

Customers rarely have a single-vendor stack of products. Therefore, it is important to provide visibility into these other products and support importing of third-party metadata that links to the information integration process. The IBM InfoSphere Information Server platform is designed to capture the essential metadata of affected business intelligence (BI) reports, data

models and sourced databases to create a comprehensive view of metadata relationships. The platform's import export manager module enables users to import the structures of BI reports, physical and glossary models from data modeling products and relational database schemas. Linkages between the IBM InfoSphere Information Server processes and third-party metadata are defined and exposed for exploration and access. This allows organizations to perform critical tasks, such as cross-tool impact analysis, to understand where change affects other areas of the integration environment.

Users can import metadata into the IBM InfoSphere Information Server repository from a wide range of third-party software tools and relational data sources (see Figure 11). For example, importing a Cognos report requires only a URL for the Cognos Content Manager and the name of the package. Importing a database schema or an XML file follows a similar procedure.

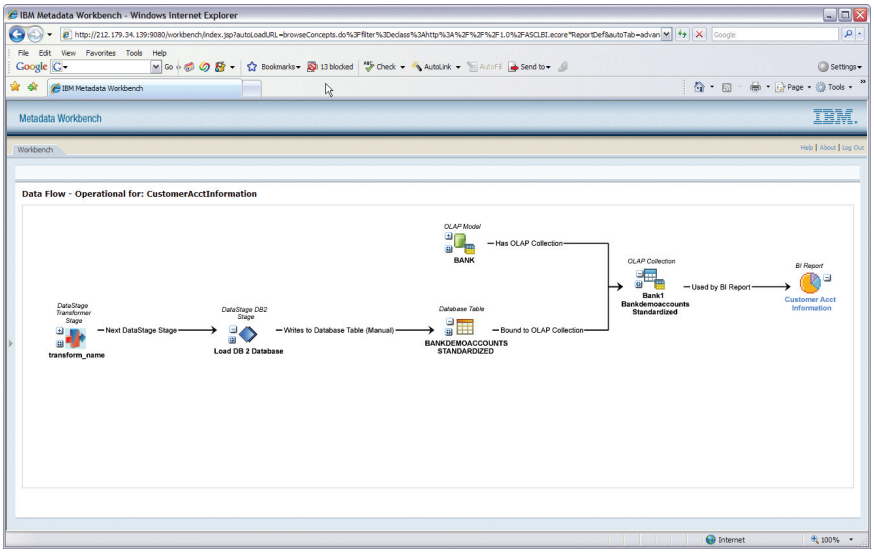
*Figure 11: Companies can use IBM InfoSphere Information Server to import third-party metadata from external business intelligence reports, data models and databases.*



### *IBM InfoSphere Metadata Workbench*

IBM InfoSphere Metadata Workbench promotes reporting, management and insight across the IBM InfoSphere Information Server modules. It provides developers and administrators with a Web-based interface to explore and navigate IBM InfoSphere Information Server metadata and third-party metadata touchpoints to reporting and modeling products (see Figure 12).

*Figure 12: IBM InfoSphere Metadata Workbench provides an intuitive, comprehensive, Web-based interface for analysis and management of information assets.*



IBM InfoSphere Metadata Workbench enables users to perform tasks in four essential categories:

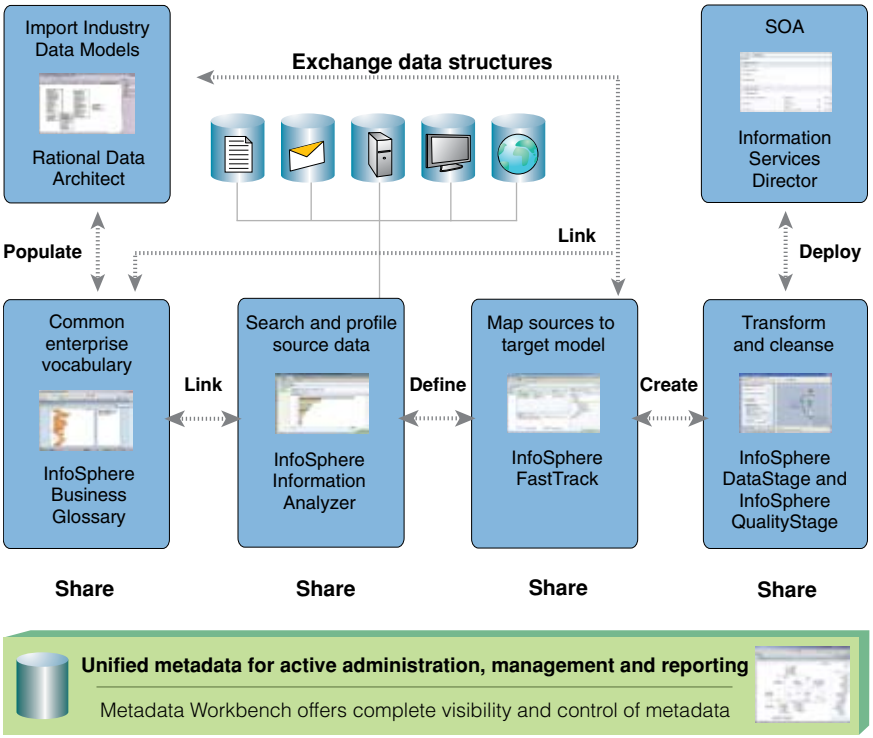
- *Explore key information assets and understand their usage, relationships and meaning; navigate through technical and business metadata assets and query information assets for simple or complex relationships to produce ad hoc reports*
- *Use IBM InfoSphere Information Server to understand information's complete lineage, including where data comes from, what it is related to and what happened to the data as it moved across applications and data warehouses*
- *Analyze the dependencies between IBM InfoSphere Information Server assets and third-party resources, such as objects from modeling and reporting tools, to perform impact analysis as well as generate reports to support compliance and governance/ regulatory standards such as the Sarbanes-Oxley Act and Basel II.*
- *Manage information assets to enable better understanding and analysis by assigning meanings and business definitions or defining relationships to new data sources*

The search and query capabilities of IBM InfoSphere Metadata Workbench are designed to enable developers, administrators, managers and analysts to view, understand and explore metadata across IBM InfoSphere Information Server to modeling and BI tools. This unique visibility promotes understanding and re-use on new projects, and can ultimately result in less duplicate development effort, shorter development times and enhanced efficiency.

### IBM InfoSphere Information Server deployment exploits unified metadata architecture

IBM InfoSphere Information Server is designed to flexibly integrate with existing organizational data integration processes. Figure 13 illustrates one possible deployment option for the platform and shows how an organization can maximize application development activities via the IBM InfoSphere Information Server unified metadata architecture.

*Figure 13: An organization can leverage the IBM InfoSphere Information Server architecture to maximize application development activities via the unified metadata repository.*



The process starts with defining data models. An organization can import information from IBM Industry Data Models (available in Rational Data Architect), which include a glossary, logical and physical data model. The glossary models contains thousands of industry-standard terms that can be used to pre-populate InfoSphere Business Glossary. Organizations can modify and extend the IBM Industry Data Models to match their particular business requirements.

After the data models are defined and business context is applied, analysts profile and understand the source systems that will be used to populate the new target data model. During the profiling process, analysts can also create and define additional new business terms as needed to describe the data sources—if these business definitions were not previously defined by the IBM Industry Data Models.

The analyst is now ready to create the mapping specifications, which are input into the ETL jobs for the new application. Using the business context and profiling results, the analyst defines the specific transformation rules necessary to convert the data sources into the correct format for the IBM Industry Data Model target. During this process, the analyst not only defines the specific business transformation rules, but also can define the direct relationship

between the business terms and their representation in physical structures. These relationships can then be published to InfoSphere Business Glossary for consumption and to enable better understanding of the asset relationships.

The business specification now serves as historical documentation as well as direct input into the generation of the InfoSphere DataStage ETL jobs. The defined business rules are directly included in the ETL job as either code or annotated to-do tasks for the developer to complete. Once the InfoSphere DataStage job is ready, the developer can also decide to deploy the same batch process as an SOA component using InfoSphere Information Services Director.

Throughout this process, metadata is generated and maintained as a natural consequence of using each of the IBM InfoSphere Information Server modules. The IBM InfoSphere Information Server platform shares relevant metadata with each of the user-specific roles throughout the entire integration process. Because of this unique architecture, managing the metadata requires little manual maintenance—unlike alternative solutions that use passive metadata and thus require extensive maintenance to manage the metadata and keep the information up-to-date. Only third-party metadata requires administration tasks such as defining the relationships to the IBM InfoSphere Information Server metadata objects. Administrators and developers who need to view both IBM InfoSphere Information Server and third party metadata assets can use InfoSphere Metadata Workbench to query, analyze and report on this information from the common repository.

### IBM InfoSphere Information Server helps companies reap the benefits of metadata for integration projects

By supporting a unified metadata strategy designed to promote collaboration, trust and compliance, IBM InfoSphere Information Server takes into account not just technology, but the people and processes as well. The platform's common, active metadata repository and services are designed to help organizations leverage metadata generated throughout the entire integration process to automatically maintain consistency across projects and teams. In addition, a highly flexible and scalable architecture supports a broad range of user roles without compromising security—giving companies the tools they need to deliver the right information to the right people and processes at the right time.

### For more information

To learn more about IBM Information Server and IBM integration solutions, visit **ibm.com**/software/data/ips



IBM Software Group
Route 100
Somers, NY 10589

Produced in the United States of America
June 2008


[1] Gartner. “Magic Quadrant for Application Infrastructure, Q207.” May 2007.
[2] Gartner. “Magic Quadrant for Data Integration Tools, 2007.” October 2007.
[3] InfoSizing, Inc. “Performance Benchmark Report: DataStage XE Parallel Extender.” December 2002.



***TAKE BACK CONTROL WITH*** **Information Management**

IMW14024-USEN-00