

Delivering information you can trust

December 2007



Information Management software

IBM Information Server FastTrack: The need for speed—accelerating data integration projects

Contents

- 3 *Creating a collaborative development environment***
- 5 *Optimizing data integration application development***
- 7 *Delivering enterprise-level benefits***
- 8 *Tracking business requirements to application deployment with FastTrack***
 - 9 *Starting a FastTrack project*
 - 12 *Profile source data*
 - 12 *Create mapping specifications to transform and cleanse data*
 - 14 *Complete data model*
 - 14 *Complete and deploy WebSphere DataStage job*
 - 15 *Generate reports*
- 17 *Integrating with IBM modules for greater functionality and productivity***
 - 18 *IBM WebSphere DataStage and IBM WebSphere QualityStage*
 - 18 *IBM WebSphere Information Analyzer*
 - 18 *IBM WebSphere Business Glossary*
 - 19 *IBM Rational Data Architect*
 - 19 *IBM Industry Data Models*
- 20 *Fast-tracking integration projects with IBM Information Server FastTrack***

For an enterprise integration project to meet business requirements and be completed on schedule and within the defined budget, business analysts, data modelers and developers must find ways to collaborate as effectively as possible. In particular, communication among individuals in these roles is critical for a successful outcome. Too often, however, development time is spent not on programming or designing, but on managing and translating differences in languages, skills and working methods, and synchronizing tool output. Schedules slip because of excessive time spent clarifying business requirements and translating how the supporting data is used in the creation of the source-to-target mappings. Typically much more time than anticipated is consumed by the back-and-forth communication necessary to clarify terms or revise incomplete or unclear specifications. In the end, differences and fresh starts commonly turn into false starts and introduce opportunities for misinterpretation.

In addition to causing communication and trust problems, the manual nature of the collaboration process makes it inefficient. In a typical workflow, the data modeler models information sources and targets and then sends these structures to the business analyst, who manually defines the source-to-target mappings by creating a Microsoft® Excel spreadsheet to print and send to the developer. The lack of automation and audit trails in these labor-intensive processes slows down the project, increases the potential for errors to be introduced and misses the opportunity to create artifacts that can be stored in a common repository and leveraged by other staff members during downstream processes.

To maximize team collaboration, increase automation and ensure projects are completed successfully and on time, IBM has introduced IBM® Information Server FastTrack. By creating an integrated environment that includes business analysts, developers and data modelers, FastTrack accelerates collaborative development across user roles, products, geographies and workflow. It automates efforts across multiple data integration tasks from analysis to code generation, while incorporating the business perspective and maintaining lineage and documented requirements. For enterprises of all sizes, FastTrack speeds time to value for strategic initiatives by accelerating the translation of business requirements into data integration projects that solve complex business problems.

Creating a collaborative development environment

FastTrack is a core module in the IBM Information Server data integration platform. Each IBM Information Server module leverages the IBM Information Server metadata services shared infrastructure during each of the various steps involved in building an integration application. The role of the FastTrack module is to track and automate efforts spanning multiple data integration tasks from analysis to code generation, thereby shortening the time from business requirements to solution implementation.

FastTrack simplifies and improves the communication process between the business analyst—who thinks in business logic—and the developer—who thinks in code. IBM Information Server, together with FastTrack, automates this information flow between the analyst—whose job is to translate business

requirements into a set of specifications for the IT staff to act on—and the developer whose job is to implement these requirements and produce a data integration application. Among other ways, it does this by converting an Excel spreadsheet deliverable into an IBM WebSphere® DataStage® job, which displays graphical representations of source-to-target mappings. For the developer, turning a partially completed WebSphere DataStage job that includes specific instructions for completion into a finished application takes much less time and is much more efficient than turning an Excel spreadsheet into the same application.

FastTrack also accelerates the development process by increasing the degree of automation in the workflow of business analysts. Rather than manually looking up data relationships in the metadata repository or elsewhere, the analyst can use FastTrack to automatically discover and compare column relationships, and to automate the process of creating associated object references. This intelligent discovery process recommends linkages between source and target structures based on evaluating the column names for exact, partial and lexical name matches. For example, FastTrack can recognize corporate abbreviation standards such as “account is synonymous with acct.” By allowing FastTrack to automate this process, the analyst can complete the mapping specification much more quickly—with fewer errors—and can turn the specification into an active object to be stored in the metadata repository and retrieved by other authorized users, who use it to shorten development time on their projects. In addition, automating the task—and the discovery process in particular—can result in a more complete specification on the first pass than would be created by manually mapping data sources.

The automated process creates reusable artifacts that are stored in the IBM Information Server metadata services repository and are accessible to authorized staff members, thereby enhancing the auditability and security of integration projects. Previously, project specifications were scattered across numerous different locations in various formats; with FastTrack, the audit trail is easily followed and questions about historical decisions can be answered quickly and with confidence.

Optimizing data integration application development

The key to optimizing development efforts is in the approach used to accelerate and automate the translation of business requirements into finished applications. FastTrack was created with this in mind and includes specific features designed to minimize the manual effort spent during design cycles, while still achieving maximum impact on downstream development activities. Features include:

- ***Automated IBM WebSphere DataStage job generation***—Automates parallel job generation and provides the ability to combine multiple specifications to create a single job; automatically derives an extract/transform/load (ETL) flow from business requirements
- ***Customizable spreadsheet view***—Enables users to annotate column mappings with business rules and/or transformation logic and to view detailed properties information including an expandable view of the artifacts in the IBM Information Server metadata services repository
- ***Automated discovery of relationships***—Features customizable discovery algorithms to find exact, partial and lexical matches on corresponding column names across source to target structures presented in a simple, easy-to-use format

- **Create and link IBM WebSphere Business Glossary terms to assets**—Enables analysts to create new business terms and document their relationship to corresponding physical columns as a part of the mapping process; this information is then published to the IBM Information Server metadata services common repository for sharing across development teams for a complete business view
- **Easy importation of data from Excel spreadsheets and .csv formats**—Allows users to bring in data from existing spreadsheets to start new specifications; also enables data to be exported into Excel spreadsheets and .csv formats
- **Flexible reporting capabilities**—Enables users to view details on column-level mappings, table-level mappings and lookups; manages and tracks metadata across the integration life cycle; provides the functionality to create PDF and HTML reports from the specifications that can be used for communicating with development teams and management
- **Table-level mappings**—Enables users to define joins, filters and case statements, while also providing the ability to leverage joins found during profiling with IBM WebSphere Information Analyzer
- **Lookup support**—Provides the ability to define lookups to generate the “lookup stage” in the generated integration job; these lookup definitions are searchable and designed for re-use
- **Generate Data Definition Language (DDL)**—Option to generate a target data model to be leveraged by data modelers and developers if one does not currently exist for the target application

Delivering enterprise-level benefits

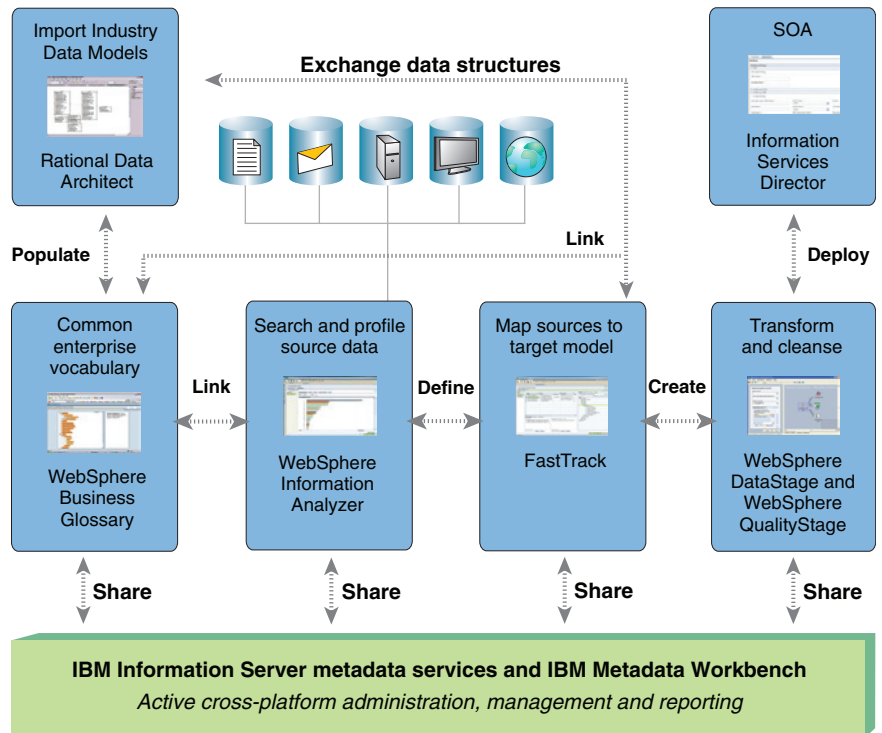
By creating a highly efficient, automated environment for developing integration applications and related design specifications, FastTrack delivers a range of business benefits, including:

- **Improved collaboration** among business analysts, developers and data architects by making it easier to consume artifacts generated by each role, thereby reducing manual efforts, misinterpretation and confusion
- **Faster time to value** for business projects by automating the supporting data integration processes
- **Improved success rates for strategic initiatives** by helping to ensure that business requirements are consistently understood and efficiently incorporated into technical solutions
- **Reduced impact of change** through the end-to-end data lineage and metadata management capabilities of the IBM Information Server platform
- **Improved data governance and auditing capabilities** by providing a centralized, managed location for project specifications and requirements
- **Improved auditability** and access to lineage between jobs and specifications

Tracking business requirements to application deployment with FastTrack

The following high-level workflow illustrates how an organization would use FastTrack as part of the IBM Information Server platform to accelerate and automate integration application development (see Figure 1). Note that security roles are maintained throughout this development process; for example, users need to be specifically assigned privileges to see or modify WebSphere Business Glossary terms or profile data sources with WebSphere Information Analyzer.

Figure 1: FastTrack fulfills a strategic role translating business requirements into executable code for application deployment



Starting a FastTrack project

Users may start a FastTrack project by importing an existing Excel spreadsheet (see Figure 2) or they can start from scratch and input the requirements

Figure 2: Users can import data from existing Excel spreadsheets or .csv formats

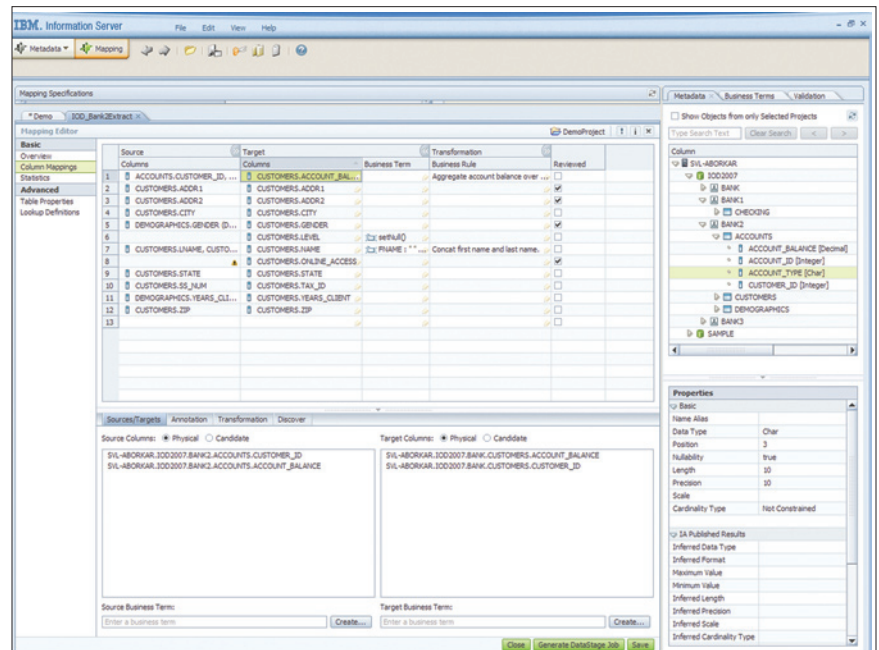
The screenshot shows a Microsoft Excel spreadsheet titled '100_BANK2Extract.xls'. The spreadsheet contains a table with the following columns: Source Columns, Target Columns, Transformation Rule, Transformation Function, and Annotation. The rows map source data from 'iod BANK2 CUSTOMERS' to target data in 'iod BANK CUSTOMERS'. For example, 'FNAME' and 'LNAME' are concatenated into a single 'NAME' field.

Source Columns	Target Columns	Transformation Rule	Transformation Function	Annotation
iod BANK2 CUSTOMERS FNAME,iod BANK2 CUSTOMERS LNAME	iod BANK CUSTOMERS NAME	Concat first name and last name.	FNAME : * * ; LNAME	
iod BANK2 CUSTOMERS ADDR1	iod BANK CUSTOMERS ADDR1			
iod BANK2 CUSTOMERS ADDR2	iod BANK CUSTOMERS ADDR2			
iod BANK2 CUSTOMERS CITY	iod BANK CUSTOMERS CITY			
iod BANK2 CUSTOMERS STATE	iod BANK CUSTOMERS STATE			
iod BANK2 CUSTOMERS ZIP	iod BANK CUSTOMERS ZIP			
iod BANK2 ACCOUNTS ACCOUNT_BALANCE,iod BANK2 ACCOUNTS CUSTOMER_ID	iod BANK CUSTOMERS ACCOUNT_BALANCE,iod BANK CUSTOMERS CUSTOMER_ID	Aggregate account balance over savings and checking accounts by customer id to compute the total account balance for each customer.		
iod BANK2 DEMOGRAPHICS YEARS_CLIENT	iod BANK CUSTOMERS YEARS_CLIENT			
iod BANK2 DEMOGRAPHICS GENDER	iod BANK CUSTOMERS GENDER			
	iod BANK CUSTOMERS ONLINE_ACCESS			
	iod BANK CUSTOMERS LEVEL		setNull()	Will be computer after standardization.

Below the spreadsheet is the 'Import from CSV' dialog box. It includes options for 'Start import from row', 'Delimiter', and checkboxes for 'Omit Hostname from Multipart Data Source Names', 'Case Sensitive Matching for Data Source Names', and 'Use First Row as Column Headers'. A preview table shows the mapping of source and target terms, transformation rules, and functions. At the bottom, there are dropdown menus for selecting source and target columns, business terms, and transformation rules, along with checkboxes for 'Join' and 'Filter'.

directly into the FastTrack interface (see Figure 3). If starting from scratch, the analyst may speed up the integration process by leveraging the IBM Industry Data Models to define the target data warehouse or application. In addition, metadata previously captured and stored in the shared IBM Information Server metadata services can be used to further accelerate the creation of source-to-target mappings, which define how data will be extracted from a source, transformed according to business rules and loaded into the target data warehouse or application. The metadata import functionality in FastTrack can also be used to import new metadata from source systems that may not already be defined within IBM Information Server.

Figure 3: FastTrack interface



Part of the core requirements for a data integration project include defining and documenting corporate standards to be applied to definitions, business transformation rules and underlying physical data structures.

Business analysts often are required to create or leverage a common set of definitions to ensure consistency in terminology across multiple development teams. Analysts may start from scratch to describe these definitions or, if corporate business terms have already been defined externally or the IBM Industry Data Models are being leveraged to accelerate project development work, this business information may be imported into WebSphere Business Glossary and leveraged directly within the FastTrack framework. Rather than just creating a common set of definitions, FastTrack enables the analyst to fully exploit existing and newly created business terms by linking each of these business terms to their physical structures to reveal a comprehensive relationship. The full lineage of business-to-technical metadata can then be published to WebSphere Business Glossary and accessed by users with the appropriate permissions.

During the mapping specification process, business transformation rules must be clearly documented to convey the business intentions to the developer in a common language. Analysts can add manual mappings as necessary to complete this critical requirement or they can leverage special functions such as discovery, which searches the metadata repository to find matches at the column level between defined sources and targets. All of this information is captured for historical purposes and presented to the developer as part of the specification process.

The business analyst is also responsible for ensuring that naming conventions are followed during the creation of these detailed mapping specifications. For example, when referring to a customer, the shortened version “cust” may be the recommended corporate standard. Physical data structure naming standards can be applied to existing or newly created data models to ensure consistency across enterprise applications.

Profile source data

Data analysts and subject matter experts (SMEs) profile data from multiple sources to understand the types of transformation rules that need to be applied during the migration process. Using WebSphere Information Analyzer and WebSphere Business Glossary content, analysts document their findings in notes, which are then shared directly with the FastTrack user and exposed in the Properties view for easy access. Often this process is iterative, so the ability to share this metadata actively via IBM Information Server metadata services is critical to streamlining this process and ensuring that details are not overlooked.

Create mapping specifications to transform and cleanse data

FastTrack gives the business analyst a spreadsheet-metaphor interface to enter source-to-target mapping specifications. These specifications can, for example, describe how to extract information from sources, filter data, map source columns to business terms and use business terms to generate a new or updated target data model and the associated table creation definitions. These mappings can be quite complex and created at multiple levels.

analyst has the option to create and apply detailed notes and annotations to the column, which will be automatically shared with the developer as part of the business requirements via the mapping specification and generated WebSphere DataStage jobs. During the mapping process, the analyst can also define the associations of business terms to the column level so other individuals can understand the full spectrum of information from a business viewpoint to its technical implementation.

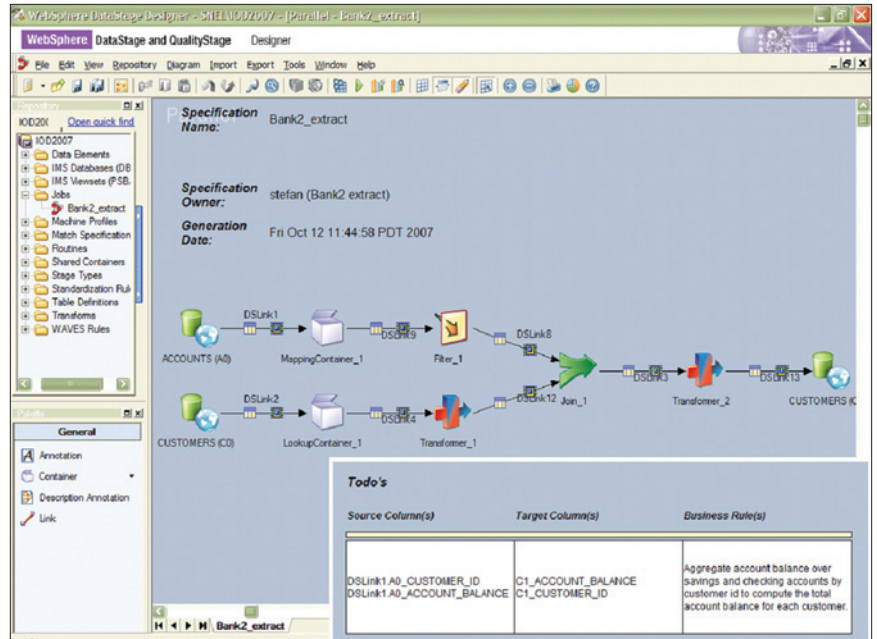
Complete data model

Once the mapping specification is complete, the updated target model structure is included in the specification details. At this point, the analyst has the option to generate a new or updated data definition language (DDL) script to pass along to the data architect and data modelers to review, enhance and deploy into the production application. Any existing naming standards and conventions can automatically be applied during this step; as a result, the data architect and data modeler may make very few changes before deploying.

Complete and deploy WebSphere DataStage job

At this point, the analyst is ready to generate WebSphere DataStage jobs to hand off to the developer to review and complete for production deployment. With a click of a mouse, the analyst transforms the business logic from one or more FastTrack specifications into a parallel WebSphere DataStage ETL job representing the data flow and rules graphically (see Figure 5). The generated job includes all transformation rules entered by the analyst so the captured business logic flows automatically through to implementation without manual intervention. These jobs can often be compiled immediately and run as-is; or,

Figure 5: FastTrack-generated WebSphere DataStage job



if additional logic or tuning is required prior to deployment of a production application, the ETL developer can edit the job and review any relevant analyst notes that are directly displayed as annotations on the job canvas.

Generate reports

FastTrack users can generate a variety of reports, including an end-to-end data transformation report with all defined specifications that can be reviewed by other project team members. The IBM Information Server core services reporting framework, a Web-based console where all platform reports are

published, generates the reports. Information about tables, table-level mapping, column-level mapping and annotations can all be added to reports, which can be published as PDF or HTML documents (see Figure 6).

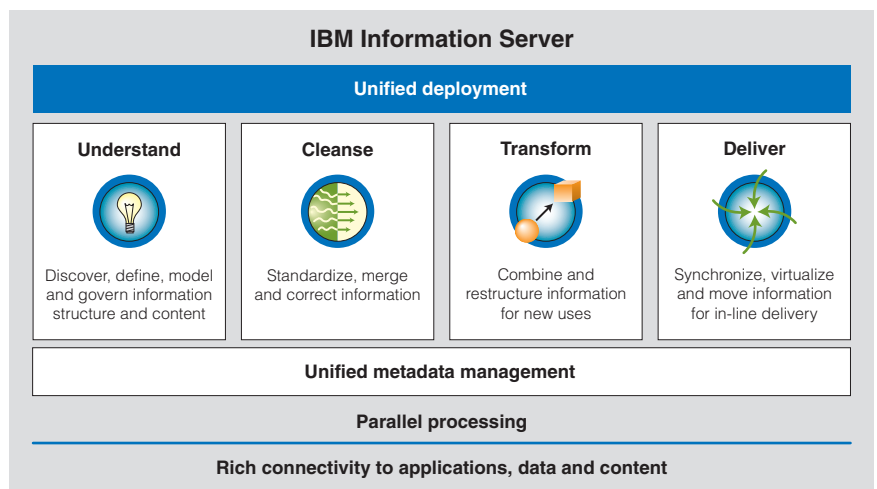
Figure 6: End-to-end report includes details on both column- and table-level mappings and lookups

Mapping Specification	
Implementation Details	
Project :	IOD
Report Name :	Bank2Extract_details
Report Generated :	2007-10-11 07:39:26
User :	admin admin
Description :	mapping specification, table mapping, column mapping, and lookup details
Specification Name :	IOD_BankExtract2_Finished
Owner :	
Status :	In Progress
Mapping Specification Description :	
Details	
Source Table :	iod.iod.BANK2.ACCOUNTS iod.iod.BANK2.CUSTOMERS iod.iod.BANK2.DEMOGRAPHICS
Target Table :	iod.iod.BANK.CUSTOMERS
Filter :	ACCOUNT_BALANCE <>-999999
Join :	iod.iod.BANK2.CUSTOMERS.CUSTOMER_ID = iod.iod.BANK2.CUSTOMERS.CUSTOMER_ID
Case :	
Reviewed :	false
Documentation :	
Details	
Source Column :	iod.iod.BANK2.ACCOUNTS.ACCOUNT_BALANCE iod.iod.BANK2.ACCOUNTS.CUSTOMER_ID
Source Term :	
Target Column :	iod.iod.BANK.CUSTOMERS.ACCOUNT_BALANCE iod.iod.BANK.CUSTOMERS.CUSTOMER_ID
Target Term :	
Transformation :	
Business Rule :	Aggregate account balance over savings and checking accounts by customer id to compute the total account balance for each customer.
Reviewed :	false
Documentation :	
All Rights Reserved. IBM is a trademark of IBM Corp. and IBM Information Server is a trademark or registered trademark in the United States, other countries or both.	

Integrating with IBM modules for greater functionality and productivity

IBM Information Server is an integrated platform built upon the IBM Information Server metadata services shared repository. It includes specific modules for every step in the process of building a data integration application (see Figure 7). FastTrack provides the most value to the development process when used in conjunction with the rest of the IBM Information Server data integration platform. To incorporate FastTrack into the development process, an enterprise should leverage IBM Information Server as a standard for the data integration process. FastTrack auto-generates IBM WebSphere DataStage jobs designed to meet the complex ETL processing demands that many organizations require as a fundamental component of their integration architecture. Other modules are optional; however, adding the modules described here can increase the automation, efficiency and capabilities of the overall development process.

Figure 7: IBM Information Server—a data integration platform designed to deliver information you can trust for key business initiatives



IBM WebSphere DataStage and IBM WebSphere QualityStage. To leverage the full value of FastTrack, users can convert their job specifications directly into ETL jobs deployed via WebSphere DataStage, an IBM Information Server module that visually transforms and aggregates large volumes of data into the proper consumable format for application access. It offers codeless, top-down visual design of data flows with hundreds of built-in transformation functions, including advanced cleansing techniques using IBM WebSphere QualityStage™. Built natively upon a massively parallel transformation engine, WebSphere DataStage supports batch, real-time and Service Oriented Architecture (SOA) processing requirements.

IBM WebSphere Information Analyzer. Providing key understanding of the content, quality and structure of source data, WebSphere Information Analyzer delivers column analysis based on actual data values (not metadata values); determines the true physical characteristics of the data, such as data type, precision, scale, nullity and others; calculates the frequency distribution; identifies the distinct values; and establishes a sample data file. Primary key, foreign key and redundant domain analysis is also part of the profiling process. WebSphere Information Analyzer replaces the time-consuming, error-prone process of performing traditional data profiling analysis by hand.

IBM WebSphere Business Glossary. WebSphere Business Glossary is a Web-based tool that enables business analysts and subject matter experts to create, manage and share a common enterprise vocabulary and classification system. Actively connected to IBM Information Server metadata services, WebSphere Business Glossary enables data stewards to link business terms to technical artifacts, which in turn are shared with other IBM Information Server modules. The result is a common set of semantic tags used by data modelers, data analysts, business analysts and end users.

IBM Rational® Data Architect. Rational Data Architect (RDA) is an enterprise data modeling and integration design tool offering logical, physical and glossary modeling. RDA is designed to help data architects understand information assets and model their relationships, design federated databases and streamline database projects. Physical and glossary information can be exchanged with IBM Information Server metadata services for use with data integration projects.

IBM Industry Data Models. The IBM Industry Data Models can be used by both business and IT to implement key strategic business initiatives faster, more reliably and with added confidence. Based upon industry experience with more than 400 clients and over 10 years of development, the IBM Industry Data Models are unique in supporting all major industry verticals including:

- *IBM Banking Data Warehouse (BDW)*
- *IBM Insurance Information Warehouse (IIW)*
- *IBM Retail Data Warehouse (RDW)*
- *IBM Telecommunications Data Warehouse (TDW)*
- *IBM Health Plan Data Warehouse (HPDW)*
- *IBM Financial Markets Data Warehouse (FMDW)*

The data models include a glossary of terms and concepts that can be clearly understood, communicated and shared between business and IT users via IBM WebSphere Business Glossary, thereby helping to accelerate project scoping, produce appropriate reports and define data requirements as well as sources of data.



Fast-tracking integration projects with IBM Information Server FastTrack

For the complex integration projects undertaken today by enterprises of all sizes, FastTrack shortens the time required to go from requirements to implementation. It does this by facilitating and improving collaboration among different users in multiple locations with different skill sets, and by automating tasks previously performed manually. As part of the integration platform available to IBM Information Server users, FastTrack helps organizations rapidly develop integrated applications that meet critical business requirements and deliver a competitive advantage.

For more information

For more information about IBM Information Server and FastTrack, contact your IBM representative or visit ibm.com/software/data/ips

© Copyright IBM Corporation 2007

IBM Software Group
Route 100
Somers, NY 10589

Produced in the United States of America
December 2007
All Rights Reserved

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

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

TAKE BACK CONTROL WITH **Information Management**