

Information Management

IBM InfoSphere Information Server: Cleansing data and managing data quality

Highlights

- ***Matches customer, vendor, product and location data based on your business rules, providing an accurate view across the enterprise***
- ***Enforces standardization, matching and survivorship rules for your core business entities***
- ***Supports multiple deployment models—executed in batch or via an application programming interface (API), or consumed as a Web service***
- ***Processes global data on a massively scalable parallel platform for optimal performance in the most demanding environments***
- ***Makes creation and maintenance of high-quality master data a reality, driving benefits across critical enterprise initiatives***

Get the most out of your organization's information assets

Organizations need to make sense of the mountains of information in their operational systems. A clear understanding of customers, partners and suppliers can mean the difference between growing a business and failing to compete. Critical initiatives for data governance, compliance and master data integration simply will not succeed unless the quality of the data in systems is clearly understood and managed. Now more than ever, business decisions must be informed — and the quality of that information is a critical element in effective decision making.

Data quality is a core discipline of IBM® InfoSphere Information Server because it enables the delivery of consistent, reusable, trusted information. The data quality continuum includes data profiling, standardization, matching and monitoring—and IBM InfoSphere Information Server delivers them all.

For more information on data profiling, visit ibm.com/software/data/infosphere/information-analyzer

Easy design for customized matching

Organizations need to ensure that strategic systems deliver accurate, complete information that business users across the enterprise can trust. Through a new, easy-to-use quality and match design user interface that can be customized to your organization's business rules, business users now have control over international names and addresses and related data such as phone numbers, birth dates, e-mail addresses and other descriptive comment fields. IBM InfoSphere Information Server matches the data elements and discovers relationships among them—in enterprise and Internet environments, and in batch and real time.

From disparate source data to high-quality information about your core business entities

IBM InfoSphere Information Server analyzes data at the character level and uncovers anomalies and buried data prior to transforming it for database loading or transaction processing. Data from disparate sources is standardized into fixed fields using business-driven rules to assign the correct semantic meaning to input data in order to facilitate matching.

Then the powerful matching capabilities of IBM InfoSphere Information Server detect duplication and relationships in the data, despite anomalous, inconsistent and missing data values. A unique statistical matching engine assesses the probability that two or more sets of data values refer to the same business entity—providing the most accurate match results available. These capabilities are delivered in an integrated design environment with transformation technology—embedding data quality into critical information integration processes.

Once a match is confirmed, IBM InfoSphere Information Server constructs linking keys so users can complete a transaction or load a target system with true entity integrity and view related data as information. As a result, by using the data quality capabilities of IBM InfoSphere Information Server in initial loads, system updates and real-time data input, companies gain access to accurate, consistent, consolidated views of any individual or business entity and its relationships across the enterprise. This powerful matching and data cleansing happens within a scalable parallel processing framework—providing unmatched performance designed for the requirements of extended enterprises.

Figures 1 and 2 show examples of how IBM InfoSphere Information Server can help standardize and transform data.

Standardization parts							
Input file:							
<i>Operation Work Instruction</i>							
WING ASSY DRILL 4 HOLE USE 5J868A HEXBOLT 1/4 INCH							
WING ASSEMBLY, USE 5J868-A HEX BOLT .25"- DRILL FOUR HOLES							
USE 4 5J868A BOLTS (HEX .25) - DRILL HOLES FOR EACH ON WING ASSEM							
RUDER, TAP 6 WHOLES, SECURE W/KL2301 RIVETS (10 CM)							
Result file:							
Assembly	Instruction	Qty	Type	Part	Size	Measure	SKU
WING	DRILL	4	HOLES	HEXBOLT	.25	INCH	5J868A
WING	DRILL	4	HOLES	HEXBOLT	.25	INCH	5J868A
WING	DRILL	4	HOLES	HEXBOLT	.25	INCH	5J868A
RUDDER	DRILL	6	HOLES	RIVET	10	CM	KL2301

Figure 1: An example of parts standardization

IBM InfoSphere Information Server data quality benefits

IBM InfoSphere Information Server provides several significant benefits as a result of enhanced data quality:

- *Helps to reduce time, cost and risk of implementing enterprise resource planning (ERP), customer relationship management (CRM), data warehouse/business intelligence and other strategic IT initiatives, and helps maximize ROI by enhancing data quality*
- *Constructs consolidated customer and household views, enabling cross-selling, up-selling and customer retention*
- *Improves customer support and service and identifies a company's most profitable customers*
- *Provides business intelligence on individuals and organizations for research, fraud detection and planning*
- *Constructs consolidated views of suppliers, parts and products for more efficient procurement and inventory management*

IBM InfoSphere Information Server data quality features

Some important features of IBM InfoSphere Information Server include:

- *An easy-to-use, integrated user interface with an intuitive, point-and-click interface for specifying automated data quality processes: data investigation, standardization, matching and survivorship*
- *Service Oriented Architecture (SOA) for creation of data quality services for real-time deployment*
- *Full integration with other IBM InfoSphere Information Server capabilities including data profiling and transformation*
- *The most powerful, accurate matching available, based on probabilistic matching technology and a full spectrum of fuzzy matching capabilities that are easy to set up and maintain*
- *Rigorous, scientific justification; easy auditing and validation*
- *Linear scalability for processing massive data volumes*

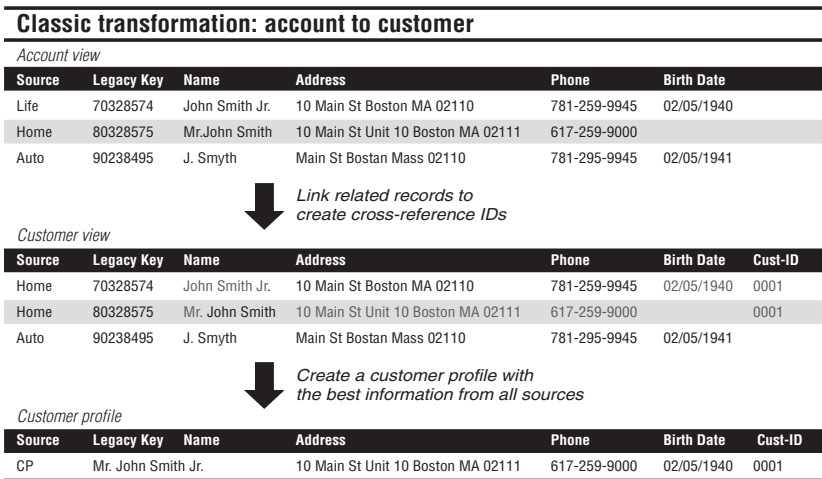


Figure 2: An example of data transformation

Data quality within a unified platform

IBM InfoSphere® QualityStage™ for IBM InfoSphere Information Server delivers these important data quality functions within the context of a complete information integration platform. It leverages unified installation and deployment for rapid startup and unified data quality and transformation functions—in combination with IBM InfoSphere DataStage® for IBM InfoSphere Information Server—reducing the number of stages and the development time for integration projects.

Active shared metadata across IBM InfoSphere Information Server simplifies the collection and management of metadata across the entire integration spectrum. Metadata definitions from IBM InfoSphere Information Analyzer can be shared and leveraged within the InfoSphere QualityStage

module—with a single import. This can result in significant benefits, including confidence in the consistency of information and the ability to perform impact analysis across IBM InfoSphere Information Server.

System requirements for InfoSphere QualityStage for IBM InfoSphere Information Server

InfoSphere QualityStage supports IBM AIX®, IBM MVS®, Linux®, Solaris, HP-UX and Microsoft® Windows® operating systems. For current, detailed hardware and software system requirements for IBM InfoSphere Information Server and other IBM Information Integration products, visit

ibm.com/software/data/integration

IBM InfoSphere Information Server delivers value

Organizations face an information challenge. Where is it? How do I get it

when I need it, in the form I need? What does it mean? What insight can I gain from it? Can I trust it? How do I control it? The challenges continue to grow if businesses cannot ensure that they have access to authoritative, consistent, timely and complete information.

IBM InfoSphere Information Server is a revolutionary new software platform that helps you derive more value from the complex, heterogeneous information spread across your systems. It enables your organization to integrate disparate data and deliver trusted information wherever and whenever needed, in line and in context, to specific people, applications and processes. It helps business and IT personnel collaborate to understand the meaning, structure and content of any type of information across any sources. It provides breakthrough productivity and performance for cleansing, transforming and moving this information consistently and securely throughout the enterprise, so it can be accessed and used in new ways to drive innovation, increase operational efficiency and help lower risk.

For more information

To learn more about IBM InfoSphere Information Server or other IBM Information Integration Solutions, contact your IBM marketing representative or IBM Business Partner, or visit

ibm.com/software/data/integration



© Copyright IBM Corporation 2009

IBM Software Group

Route 100
Somers, NY 10589
U.S.A.

Printed in the United States of America
August 2009
All Rights Reserved

IBM, the IBM logo, ibm.com, AIX, DataStage, MVS, and InfoSphere are trademarks of International Business Machines Corp., registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the Web at "Copyright and trademark information" at ibm.com/legal/copytrade.shtml

Linux is a registered trademark of Linus Torvalds in the United States, other countries or both.

Microsoft, and Windows are trademarks 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.

IMD11784-USEN-01