



---

## Highlights

- Investigates data to identify the as-is level of data quality and determine quality issues
  - Enforces standardization, matching and data survivorship rules for core business entities
  - Matches customer, vendor, product and location data based on an organization's business rules, enabling an accurate, consistent view across the enterprise
  - Processes global data on a massively scalable parallel platform for optimal performance
  - Delivers reliable, high-quality data to critical enterprise initiatives to enable success in both batch architectures and Service Oriented Architectures (SOAs)
- 

# IBM InfoSphere QualityStage

*Investigate, cleanse and manage high-quality data to deliver better business results*

## 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, products, partners and suppliers makes the difference between growing a business and failing to compete. Without clean, standardized and accurate data, that clear understanding cannot be achieved. In turn, poor data quality contaminates and undermines critical business initiatives, such as information governance, compliance and master data management.

Most organizations, however, have not yet evolved their processes, policies and infrastructure to enable high data quality. To address this need, organizations are increasingly adopting information governance, a quality-control discipline that adds new rigor to the process of defining common terminology and managing, using, improving and protecting information. Effective information governance can enhance the quality, availability and integrity of a company's data by fostering cross-organizational collaboration and structured policy making.

Data quality is a key part of information governance and is a core discipline within the IBM® InfoSphere™ Information Server platform, helping to enable the delivery of consistent, accurate, trusted information. With the InfoSphere Information Server data integration platform, IBM delivers a wide range of data quality capabilities, from data profiling, standardization and matching to active data quality monitoring.



Organizations focus on different aspects of data quality at different points in time; InfoSphere Information Server provides several capabilities that address data quality needs for each of those touch points.

IBM InfoSphere QualityStage™, part of the InfoSphere Information Server data integration platform, focuses on cleansing data: it enables enterprises to create and maintain accurate views of key entities, including customers, vendors, locations and products. Core InfoSphere QualityStage capabilities include data investigation, standardization, address verification, probabilistic matching, data survivorship and data enrichment. InfoSphere QualityStage may be deployed in transactional, operational or analytic environments, in batch or in real time.

IBM InfoSphere Information Analyzer, also part of the InfoSphere Information Server platform, delivers another set of data quality enhancement capabilities to help clients understand, analyze and monitor data. With integrated rules analysis, exception management and an intuitive user interface, clients can maintain high-quality data to help achieve business objectives. For more information about InfoSphere Information Analyzer, visit: [ibm.com/software/data/integration/information-analyzer](http://ibm.com/software/data/integration/information-analyzer)

**InfoSphere QualityStage:  
A path to data quality benefits**

InfoSphere QualityStage is designed to deliver high-quality data and help organizations reap related benefits, including:

- Improved return on investment (ROI)
- Reduced time, cost and risk of implementing enterprise resource planning (ERP), customer relationship management (CRM), data warehousing, business intelligence, master data management and other strategic IT initiatives
- Cleansed, consolidated customer and household views that support cross-selling and up-selling efforts
- Improved customer support and service, with the ability to identify the most profitable customers
- Consolidated views of suppliers, parts and products for more efficient analysis, procurement and inventory management
- Tight integration with the broader InfoSphere Information Server data integration platform, enabling a holistic approach that makes data quality a key component of data integration

Figures 1 and 2 show examples of how InfoSphere QualityStage can help standardize and transform data.

---

**Standardization parts**

---

**Input file:**

*Operation Work Instruction*  
 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 product parts standardization

**Classic transformation: account to customer**

*Account view*

Source	Legacy Key	Name	Address	Phone	Birth Date	Cust-ID
Life	70328574	John Smith Jr.	10 Main St Boston MA 02110	781-259-9945	02/05/1940	
Home	80328575	Mr. John Smith	10 Main St Unit 10 Boston MA 02111	617-259-9000		
Auto	90238495	J. Smyth	Main St Bostan Mass 02110	781-295-9945	02/05/1941	

*Customer view*

↓ **Link related records to create cross-reference IDs**

Source	Legacy Key	Name	Address	Phone	Birth Date	Cust-ID
Life	70328574	<a href="#">John Smith Jr.</a>	10 Main St Boston MA 02110	781-259-9945	<a href="#">02/05/1940</a>	<a href="#">0001</a>
Home	80328575	<a href="#">Mr. John Smith</a>	<a href="#">10 Main St Unit 10 Boston MA 02111</a>	<a href="#">617-259-9000</a>		<a href="#">0001</a>
Auto	90238495	J. Smyth	Main St Bostan Mass 02110	781-295-9945	02/05/1941	<a href="#">0002</a>

*Customer profile*

↓ **Create a customer profile with the best information from all sources**

Source	Legacy Key	Name	Address	Phone	Birth Date	Cust-ID
CP		Mr. John Smith Jr.	10 Main St Unit 10 Boston MA 02111	617-259-9000	02/05/1940	0001
CP		J. Smyth	Main St Bostan Mass 02110	781-295-9945	02/05/1941	<a href="#">0002</a>

Figure 2. An example of data transformation

Organizations must ensure that strategic systems deliver accurate, comprehensive information that business users across the enterprise can trust. Through its easy-to-use, customizable user interface, InfoSphere QualityStage helps business users gain control over international names and addresses, and related data such as phone numbers, birth dates, email addresses and other descriptive comment fields. InfoSphere QualityStage uses highly accurate probabilistic matching algorithms to match data elements and discover relationships among them—in enterprise and Internet environments, and for batch and real-time processing.

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

By performing character-level analysis, InfoSphere QualityStage helps uncover anomalous and buried data prior to transforming it for database loading or transaction processing. First, data from disparate sources is standardized into fixed fields, and business-driven rules assign the correct semantic meaning to the input data in order to facilitate matching.

Next, the powerful matching capabilities of InfoSphere QualityStage detect duplication and relationships in the data, despite anomalous, inconsistent or 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 extremely accurate match results. These capabilities are delivered in an integrated design environment with transformation technology, which helps embed data quality into critical information integration processes.

Once a match is confirmed, InfoSphere QualityStage constructs linking keys so users can complete a transaction or load a target system with true entity integrity, and can view related data as information. By using the data quality enhancement capabilities of InfoSphere Information Server during initial loads and system updates and during 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 occurs within a scalable parallel processing framework—providing world-class performance designed for the requirements of extended enterprises.

### InfoSphere QualityStage features

- Easy-to-use, integrated and intuitive point-and-click user interface for specifying automated data quality processes: data investigation, standardization, matching and survivorship
  - Enhanced Match Designer tool that enables easier setup and greater flexibility
  - Global address cleansing, validation, certification (for specific localities) and geolocation
  - Standardization and match reporting to gain greater insight into your data quality process and improve the quality of deployments
  - Additional standardization rules to set coverage for Latin America, the Netherlands and India, as well as coverage for traditional Chinese and Japanese kana
  - Rules-set acceleration for product data
  - SOA for creation of data quality services for real-time deployment
  - Powerful, accurate matching based on probabilistic matching technology and a full spectrum of fuzzy matching capabilities that are easy to set up and maintain
  - Rigorous, scientific justification of matching, plus easy auditing and validation
  - Efficient runtime and system resource usage and massive scalability
  - Full integration with other InfoSphere Information Server capabilities including shared metadata, data monitoring, profiling and transformation
- 

### Data quality within a unified platform

As part of the InfoSphere Information Server platform, InfoSphere QualityStage delivers important data quality functions within the context of a complete information integration platform. It leverages unified installation, deployment and source control for rapid startup as well as unified data quality and transformation functions—in combination with IBM InfoSphere DataStage®—to help reduce the development time for integration projects and help ensure the quality of delivered data.

Active shared metadata across the InfoSphere Information Server platform helps simplify the collection and management of metadata over the entire integration spectrum. Metadata from InfoSphere Information Analyzer can be shared and leveraged within InfoSphere QualityStage, enabling superior collaboration. This level of integration can result in significant benefits, including greater confidence in the consistency of information and the ability to perform impact analysis across InfoSphere Information Server.

### Data quality and information governance

Information governance can enhance the quality, availability and integrity of a company's data and foster cross-organizational collaboration and structured policy making. Applied consistently, it can help balance factional silos with organizational interest, directly impacting four of the most important objectives of any business: increasing revenue, lowering costs, reducing risks and increasing confidence in its data. Additionally, information governance allows an organization to monitor its information supply chain as an end-to-end system, helping to ensure that information is consistently defined and well understood; reliable and of high quality; managed throughout its life cycle; and protected wherever it lies.

InfoSphere Information Server, InfoSphere QualityStage and InfoSphere Information Analyzer deliver the data quality functionality organizations need to institute and enable information governance policies.

### A forum for information governance

Now more than ever, data protection and management is a universal business concern. To help organizations better understand the emerging information governance field, IBM created a leadership forum in November 2004 for chief data officers and security, risk, compliance and privacy officers concerned about information governance issues.

Since then, the IBM Information Governance Council has steadily grown to comprise nearly 55 leading companies, universities and IBM Business Partners, including large financial institutions, telecommunications organizations, retailers and government agencies. The Council designed a framework to help businesses understand the core and supporting disciplines and the enablers of information governance. It also produced a maturity model to help assess information governance within an organization. To broaden involvement in the Council, IBM launched an online community to encourage organizations to participate, using crowdsourcing technology to further enhance the maturity model and information governance as a whole.

For more information on the IBM Information Governance Council, please visit: [www.infogovcommunity.com](http://www.infogovcommunity.com)

---

### InfoSphere Information Server delivers value

Organizations face ongoing challenges with information: Where is it? How do I get it when I need it, in the form I need? Can I trust it? How do I control it? The hurdles continue to mount if businesses cannot ensure that they have access to authoritative, consistent, timely and complete information.

InfoSphere Information Server is a market-leading data integration platform that helps organizations derive more value from the complex, heterogeneous information spread across their systems. It enables an 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 range of 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 InfoSphere QualityStage, including detailed hardware and software system requirements, please contact your IBM marketing representative or IBM Business Partner, or visit: [ibm.com/software/data/infosphere/qualitystage](http://ibm.com/software/data/infosphere/qualitystage)

For more information about data quality solutions from IBM, visit: [ibm.com/software/data/integration/capabilities/cleanse.html](http://ibm.com/software/data/integration/capabilities/cleanse.html)

To learn more about InfoSphere Information Server or other IBM information integration solutions, please contact your IBM marketing representative or IBM Business Partner, or visit: [ibm.com/software/data/integration](http://ibm.com/software/data/integration)



---

© Copyright IBM Corporation 2011

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

Produced in the United States of America  
February 2011  
All Rights Reserved

IBM, the IBM logo, [ibm.com](http://ibm.com), InfoSphere and QualityStage are trademarks or registered trademarks of International Business Machines Corporation in the United States, other countries or both. If these and other IBM trademarked terms are marked on their first occurrence in this information with a trademark symbol (® or ™), these symbols indicate U.S. registered or common law trademarks owned by IBM at the time this information was published. Such trademarks may also be registered or common law trademarks in other countries. A current list of IBM trademarks is available on the web at "Copyright and trademark information" at [ibm.com/legal/copytrade.shtml](http://ibm.com/legal/copytrade.shtml)

Other product, company 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. All statements regarding IBM's future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only.



Please Recycle