

IBM InfoSphere Information Server for Data Quality

A component index

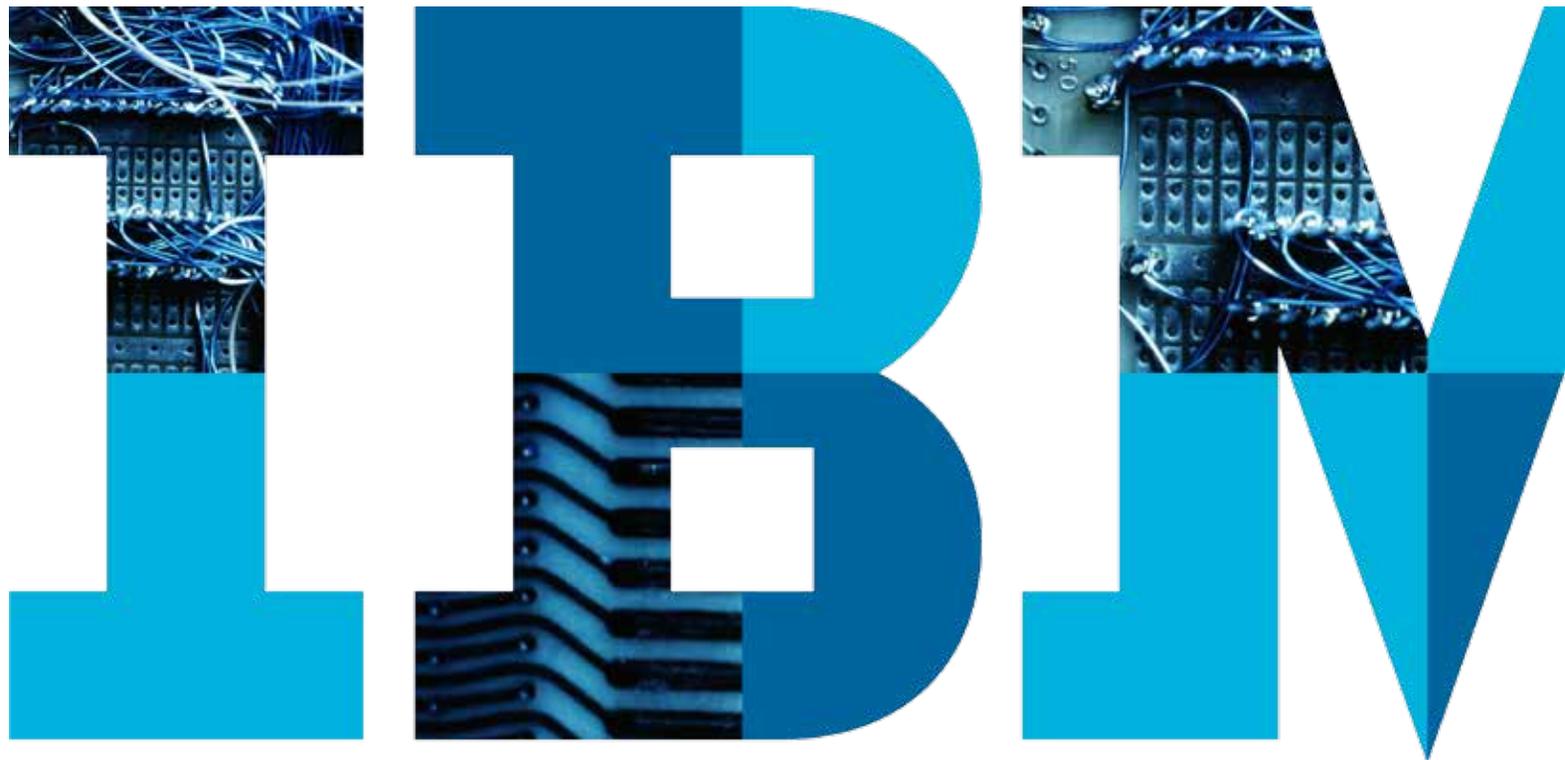


Table of contents

Information management challenges	3
A blueprint for an effective information architecture	6
Data cleansing tools	9
InfoSphere Information Server for Data Quality	9
InfoSphere QualityStage	10
InfoSphere Information Analyzer	12
InfoSphere Discovery	13
The bottom line: Gaining competitive advantage	14

Information management challenges

A blueprint for an effective information architecture

Data cleansing tools

The bottom line: Gaining competitive advantage

Information management challenges

Do you have confidence in your organization's data and in the results that data produces? Is your data organized in a coherent infrastructure, and easily accessible by the right people, processes and applications throughout the organization? Can the decision makers who need the information call it up on demand?

If you're asking these questions frequently, congratulations: you're thinking about the details that will put your business in a position to succeed. But to realize that success, each answer must be a resounding "yes"—and intelligent information management strategies can help you get there.

Data is a valuable asset that must be managed as it moves through an organization. Yet just as information sources are growing more numerous and diverse, and regulatory compliance initiatives more focused, the need to integrate and access information from these disparate sources in consistent, trusted and reusable ways is also



becoming critical. To meet this challenge, organizations need solutions that:

- Provide decision makers with a hub for actionable information (data warehousing)
- Provide the scalability and agility necessary for massive data volumes (big data analytics)
- Efficiently deliver trusted information to the analytics infrastructure (business intelligence)
- Reduce costs through hardware and software consolidation (infrastructure rationalization)
- Improve information governance practices to better manage risk control and compliance (information governance)

Information management challenges

A blueprint for an effective information architecture

Data cleansing tools

The bottom line: Gaining competitive advantage

The inability to address any one of these issues greatly reduces an organization's chances of capitalizing on new market opportunities, responding to new customer demands or counteracting competitive threats.

Confidence depends upon the quality, availability and integrity of an organization's data. Excellent data quality is achieved through several essential attributes:

- **Completeness:** All related data must be linked from all possible sources.
- **Accuracy:** Data must be correct and consistent, with common data problems remediated, such as misspellings or abbreviations.
- **Availability:** Data must be available upon demand.
- **Timeliness:** Current data must be available.

What's holding back your information integration efforts?

Several technical and operational factors can disrupt attempts to deploy information-driven solutions or adjust systems to enhance access and processes. Here are a few of the top roadblocks:

- **Multiple versions of the truth** prevent organizations from effectively complying with information-centric regulations or achieving a single view of customers, products, accounts or locations.
- **Data resides in multiple databases and applications** with little governance of consistency or accuracy.
- **Information overload** prevents workers and systems from prioritizing and differentiating data.
- **A lack of metadata**, which provides information about the data itself, prevents IT and business from collaborating on the specific meaning and usage of information and terminology.
- **A lack of trust in information** because different information sources deliver different answers to the same question.
- **An inability to efficiently standardize, merge and correct information from multiple sources** breeds disdain for inaccurate information and delays adoption of new business applications.
- **Flexibility and agility problems** due to information that is tightly coupled to specific applications and processes, preventing the natural evolution of IT architecture as systems are added, updated and retired.

Information management challenges

A blueprint for an effective information architecture

Data cleansing tools

The bottom line: Gaining competitive advantage

IBM® InfoSphere® Information Server for Data Quality offers end-to-end data quality capabilities that help organizations:

- Understand data and its relationships to gain a complete picture before beginning a project
- Analyze and monitor data quality continuously to reduce the proliferation of incorrect or inconsistent data
- Cleanse, standardize and match data to ensure its quality and consistency and to provide a single version of the truth
- Maintain data lineage so end users can trace data back to original sources, establishing trust and confidence in the information received

This e-book outlines the components of IBM InfoSphere Information Server for Data Quality, and describes how its combination of tools and expertise provides an exceptionally deep and feature-rich platform upon which your organization can create and maintain high-quality data.

Information management challenges

A blueprint for an effective information architecture

Data cleansing tools

The bottom line: Gaining competitive advantage

A blueprint for an effective information architecture

The key to creating and maintaining confidence in an organization's information is a sustainable, agile and governable enterprise information architecture. To deliver trusted information wherever, whenever and however it is needed—across a full range of business requirements—a comprehensive information integration solution must:

- Connect to relevant applications, data and content, and recognize and respond to data changes in those sources—structured or unstructured, mainframe or distributed, internal or external
- Discover, model and govern information structure and content before the data is integrated and proliferated throughout the enterprise
- Standardize, merge and correct information to provide authoritative, consistent and complete views of business information and its relationships across the extended enterprise



- Effectively and efficiently collect, combine and restructure high volumes of data for new uses
- Synchronize, virtualize and move information for in-line delivery
- Flexibly publish and manage reusable information services in a service-oriented architecture (SOA) model

Information management challenges

A blueprint for an effective information architecture

Data cleansing tools

The bottom line: Gaining competitive advantage

InfoSphere Information Server is designed to deliver these capabilities. It helps business and IT personnel collaborate to understand the meaning, structure and content of any type of information across any source. It also provides breakthrough productivity and performance for cleansing, transforming and moving this information consistently and securely throughout the enterprise so it can be used to drive innovation, increase operational efficiency and help lower risk.

The capabilities of InfoSphere Information Server are bundled into four purpose-oriented packages (see Figure 1):

- **IBM InfoSphere Business Information Exchange:** Understand data and foster collaboration between IT and line-of-business teams to narrow the communication gap and create business-driven information integration.

- **IBM InfoSphere Information Server for Data Integration:** Transform data in any style and deliver it to any system, supporting faster time-to-value and reduced risk for IT.
- **IBM InfoSphere Information Server for Data Quality:** Establish and manage high-quality data, turning a deluge of information into a trusted asset.
- **IBM InfoSphere Information Server Enterprise Edition:** Gain the capabilities of all three individual packages in one comprehensive package—you can start your information integration efforts in one area, and then be ready to expand when needed to further optimize results.

Each of these packages contains a set of tools—metadata management, data cleansing, data integration and others—to speed information integration.

Information management challenges

A blueprint for an effective information architecture

Data cleansing tools

The bottom line: Gaining competitive advantage

IBM Information Integration

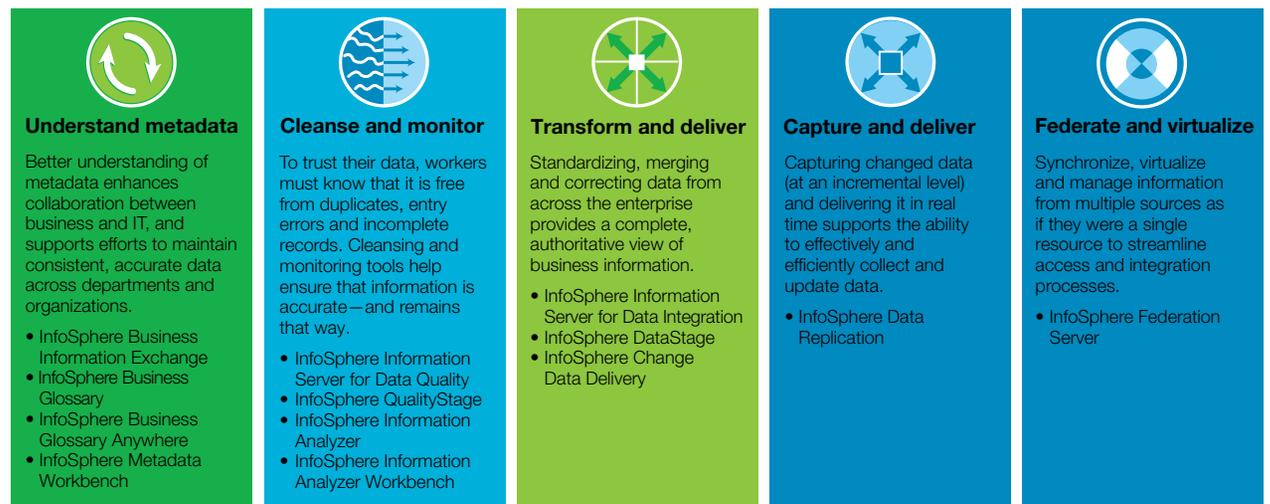


Figure 1. InfoSphere Information Server helps companies support data integration and data quality requirements while promoting collaboration between IT and business teams.

Information management challenges

A blueprint for an effective information architecture

Data cleansing tools

InfoSphere Information Server for Data Quality

InfoSphere QualityStage

InfoSphere Information Analyzer

InfoSphere Discovery

The bottom line: Gaining competitive advantage



Cleanse and monitor

To trust their data, workers must know that it is free from duplicates, entry errors and incomplete records. Cleansing and monitoring tools help ensure that information is accurate—and remains that way.

- InfoSphere Information Server for Data Quality
- InfoSphere QualityStage
- InfoSphere Information Analyzer
- InfoSphere Information Analyzer Workbench

Data cleansing tools

InfoSphere Information Server for Data Quality

InfoSphere Information Server for Data Quality provides rich capabilities for organizations to cleanse data and monitor data quality on an ongoing basis, helping turn data into trusted information that can be used to inform business decisions and streamline the execution of business processes.

The software delivers comprehensive and customizable data cleansing capabilities in batch and real time to automate source data investigation, information standardization and record matching—all based on business rules defined by the company. Additionally, organizations can enrich data and make sure the best data across sources survives.

InfoSphere Information Server for Data Quality also helps monitor and maintain data quality. By establishing data quality metrics that align with business objectives, organizations can quickly uncover data quality issues and establish a remediation plan. The following InfoSphere Information Server for Data Quality products offer capabilities that help organizations properly manage a data governance program:

- InfoSphere QualityStage®
- InfoSphere Information Analyzer
- InfoSphere Discovery

For more information about InfoSphere Information Server for Data Quality, visit: ibm.com/software/data/quality

Information management challenges

A blueprint for an effective information architecture

Data cleansing tools

InfoSphere Information Server for Data Quality

[InfoSphere QualityStage](#)

InfoSphere Information Analyzer

InfoSphere Discovery

The bottom line: Gaining competitive advantage

InfoSphere QualityStage

InfoSphere QualityStage enables enterprises to create and maintain an accurate view of master data entities, such as customers, vendors, locations and products. It provides a development environment with a flexible set of capabilities and an intuitive, “design-as-you-think” user interface. The software matches data using probabilistic algorithms designed to ensure that the information needed to run the enterprise is accurate and trustworthy. It processes global data on a massively scalable parallel platform for optimal performance in demanding environments.

InfoSphere QualityStage enables a comprehensive process to manage and maintain data quality through several core functions:

- **Investigation:** Understand the nature and extent of data anomalies and enable more effective cleansing and matching.
- **Standardization:** Create a standardized view of customer, partner or product data. This capability also enables global address cleansing, validation and certification (for significant postal discounts in select localities) and geolocation.
- **Probabilistic matching:** Provides an industry-leading matching engine to help ensure the best match results possible; built on a platform enabled for high connectivity and scalability.
- **Survivorship:** Helps ensure the optimum consolidation, householding or linked view of record information; enables a consolidated and accurate view of customers, partners, products and more.

Information management challenges

A blueprint for an effective information architecture

Data cleansing tools

InfoSphere Information Server for Data Quality

[InfoSphere QualityStage](#)

InfoSphere Information Analyzer

InfoSphere Discovery

The bottom line: Gaining competitive advantage

The probabilistic matching capability and dynamic weighting strategies of InfoSphere QualityStage help you create high-quality, accurate data. With InfoSphere QualityStage, you can consistently identify core business information such as customer, location and product throughout the enterprise. InfoSphere QualityStage standardizes and matches any type of information. By promoting data quality, InfoSphere QualityStage reduces the time and cost to implement customer relationship management (CRM), business intelligence, enterprise resource planning (ERP) and other strategic customer-related IT initiatives.



For more information about InfoSphere QualityStage, visit: ibm.com/software/products/us/en/ibminfoqual

Information management challenges

A blueprint for an effective information architecture

Data cleansing tools

InfoSphere Information Server for Data Quality
InfoSphere QualityStage
[InfoSphere Information Analyzer](#)
InfoSphere Discovery

The bottom line: Gaining competitive advantage

InfoSphere Information Analyzer

IBM InfoSphere Information Analyzer helps you scope your data quality projects, develop metrics to form a complete picture of data quality, and continuously monitor data health using an easy-to-understand dashboard. The artifacts delivered by InfoSphere Information Analyzer enable data owners to focus on detecting and responding to critical data quality issues, and to deliver trusted data to the enterprise. By creating and reusing rules across multiple data sources, the organization can improve time-to-value and deliver more consistent and correct data.

Rules analysis, a key data assessment capability, extends the ability to compare, evaluate, analyze and monitor expected data quality. It consists of data rules that evaluate data through focused and targeted testing of data against user-defined conditions.

By combining multiple rules, InfoSphere Information Analyzer provides a broad, holistic assessment of records and data sources and supports rules analysis at multiple levels. As rules are defined logically, they can be developed once and applied repeatedly and consistently to any number of data sources. For example, the rule you deploy against multiple data sources can also be applied as part of an extract, transform and load (ETL) or data-cleansing job.

For more information about InfoSphere Information Analyzer, visit: ibm.com/software/products/us/en/ibminfoanal

Information management challenges

A blueprint for an effective information architecture

Data cleansing tools

InfoSphere Information Server for Data Quality

InfoSphere QualityStage

InfoSphere Information Analyzer

[InfoSphere Discovery](#)

The bottom line: Gaining competitive advantage

InfoSphere Discovery

InfoSphere Discovery automates the discovery of data relationships—from simple to complex, within and across heterogeneous systems. This comprehensive solution allows users to create a complete, 360-degree view of their data assets. Organizations can achieve reduction in analysis time, greater accuracy and higher levels of visibility into potential data problems.

By automating the complex process of data discovery, InfoSphere Discovery replaces error-prone, manual data inspection methods and enables organizations to discover hidden data relationships to define business objects such as customer, patient or invoice. InfoSphere Discovery is also used to identify the transformation rules that have been applied to a source system when populating a target such as a data warehouse or an operational data store. Once accurately defined, these business objects and transformation rules provide the essential input into information-centric projects such as data integration, master data management (MDM) and archiving.

InfoSphere Discovery complements the IBM InfoSphere Optim™ Data Growth solution by capturing both database-defined and application-managed relationships that make up the business object to be archived. By using these products together, businesses can easily implement best-practice archiving capabilities to reduce storage requirements and lower costs. InfoSphere Discovery also enables businesses to accurately implement data consolidation projects by allowing IT staff to prototype and test new transformation rules for completeness before they convert and move any data.

For more information about InfoSphere Discovery, visit:
ibm.com/software/products/us/en/infospherediscovery

Information management challenges

A blueprint for an effective information architecture

Data cleansing tools

The bottom line: Gaining competitive advantage

The bottom line: Gaining competitive advantage

Using information effectively helps your organization outperform competitors, create innovation and build shareholder value. The value of accurate, trusted data—and the confidence it gives your workers—cannot be ignored in today’s fast-paced, always-on business climate.

Whether the focus is a single integration project or a broad infrastructure initiative, IBM InfoSphere Information Server for Data Quality provides the scalable and reliable technology you need to reduce risk and accelerate success. With IBM tools and solutions, you can analyze and cleanse information from and across diverse data sources, make information reusable and easier to share, process high volumes of information in different ways based on business demands and provide the real-time information necessary to build an agile business.





© Copyright IBM Corporation 2014

IBM Corporation
Software Group
Route 100
Somers, NY 10589

Produced in the United States of America
February 2014

IBM, the IBM logo, ibm.com, DataStage, InfoSphere, Optim, and QualityStage 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

This document is current as of the initial date of publication and may be changed by IBM at any time. Not all offerings are available in every country in which IBM operates.

THE INFORMATION IN THIS DOCUMENT IS PROVIDED “AS IS” WITHOUT ANY WARRANTY, EXPRESS OR IMPLIED, INCLUDING WITHOUT ANY WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND ANY WARRANTY OR CONDITION OF NON-INFRINGEMENT. IBM products are warranted according to the terms and conditions of the agreements under which they are provided.

The client is responsible for ensuring compliance with laws and regulations applicable to it. IBM does not provide legal advice or represent or warrant that its services or products will ensure that the client is in compliance with any law or regulation.



Please Recycle
