IBM InfoSphere
Information Analyzer

Understand, analyze and monitor your data

---

**Highlights**

- Develop a greater understanding of data source structure, content and quality
- Leverage data quality rules continuously for greater validation, trending and pattern analysis
- Identify, measure and manage exceptions to limit data quality decay and reduce risk

---

**Understanding and addressing data quality: A critical early step**

What price tag would you place on the value of truly understanding the data sources of your data-driven project? According to industry analysts, more than 80 percent of IT projects either fail or overrun their budgets due to a lack of data understanding, and low data quality can cost companies billions of dollars annually in lost productivity, inefficient service and inaccurate materials.

Data is constantly changing and shifting as it moves across your distributed data landscape. Human error and software bugs cause data errors that break business rules and quickly propagate across the company. Additionally, business rules, data relationships and transformation logic are always in flux. If your distributed data environment isn’t monitored and remediated, data consistency and quality will decay, causing discrepancies that result in poor business decisions or mistakes that violate regulatory requirements.

IBM® InfoSphere™ Information Analyzer helps you quickly and easily understand data by offering data quality assessment, data quality monitoring and flexible data rule design and analysis capabilities. Those insights enable you to derive more information from enterprise data to accelerate information-centric projects.

InfoSphere Information Analyzer is a component of IBM InfoSphere Foundation Tools, a collection of solutions that help organizations jump-start their data integration efforts. InfoSphere Foundation Tools combine data discovery and understanding, data modeling and mapping, business rule specification creation, data stewardship, business vocabulary management, information lineage and metadata management, all with a shared repository. Data profiling and analysis play an integral role in both InfoSphere Information Analyzer and InfoSphere Foundation Tools, providing key insight into the structure, content and quality of trusted information.
Data quality is a core discipline of InfoSphere Information Analyzer because it enables the delivery of consistent, reusable, trusted information. The data quality continuum includes data profiling, definition and testing of validation and transformation rules and data quality monitoring and remediation. InfoSphere Information Analyzer works in conjunction with IBM InfoSphere QualityStage™ software for standardization and matching. For more information on InfoSphere QualityStage, visit ibm.com/software/data/integration/qualitystage

**InfoSphere Information Analyzer inspires confidence in data quality**

The InfoSphere Information Analyzer Investigate module provides data profiling as well as domain and cross-domain analysis; the Develop module supports the creation, evaluation and trending of data quality rules and metrics. InfoSphere Information Analyzer Exception Manager provides continuous data monitoring and exception management of business rules and transformations discovered and those already known, such as transformations from a mapping specification or captured in IBM InfoSphere Business Glossary. This helps not only to simplify information integration but also to inspire greater confidence in the quality of information delivered.

InfoSphere Information Analyzer allows you to meet the demanding requirements of your enterprise today while achieving the data quality objectives that will help sustain your growth for tomorrow. InfoSphere Information Analyzer can help you:

- Expedite delivery of data-driven projects
- Reduce the risk and impact of proliferating incorrect and inaccurate data
- Utilize business-driven Rules Analysis (see Figure 1) with reusable construction and application across multiple data sources to offer quick time to value
- Minimize costs and resources necessary for critical data integration projects
- Facilitate the timely delivery of trusted information

*Figure 1*: InfoSphere Information Analyzer can display graphical views of Rules Analysis quality results and trends.
InfoSphere Information Analyzer profiles data by performing complete source-system analysis, including column analysis, primary and foreign key analysis and cross-domain relationship analysis on your heterogeneous data sources. It also features extensive reporting capabilities, robust security, scheduling features and a customizable dashboard, all in a state-of-the-art intuitive GUI.

Rules Analysis is a key data quality assessment capability that extends your ability to compare, evaluate and analyze data. It consists of data rules that evaluate data through focused and targeted testing of data against user-defined conditions. By combining multiple rules, it can provide a broad, holistic assessment of records and data sources, enabling rules analysis at multiple levels.

InfoSphere Information Analyzer data quality assessment features include:

**Comprehensive data analysis**—A complete set of metrics offers a holistic picture of your data from many angles; enables users to immediately document all discovered data anomalies including data type, minimum, maximum, count, nulls and averages; and enables users to identify incomplete and invalid values (see Figure 2).

**Real-time drill-down into records**—Individual records from profiling results can be viewed in real time. For example, if you discover an invalid value in a column, you can easily drill down to the actual record for further investigation.

**Integrated Rules Analysis**—This robust capability provides development, deployment and evaluation of critical data rules on an ongoing basis (see Figure 3). It features holistic, multilevel rule assessment at the rule, record and source levels for greater insight into potential quality issues. Rules can be built free-form or through a structured builder, helping users to readily compose standard data conditions.

**Scheduled deployments**—As rules are defined logically, they can be developed once and applied repeatedly and consistently to any number of data sources. The resulting data rules can be run in ad hoc or scheduled modes, or deployed into production environments for ongoing data quality monitoring.

**Intuitive user interface and dashboard**—A rich, graphically enabled user interface with intuitive navigation experience allows you to quickly gain an overall quality assessment and begin evaluating concerns. A task-based approach guides you toward resolution of the most serious concerns.
Assessment of data quality trends—Results of rules or more comprehensive rule sets can be measured against established benchmarks or thresholds. Additional metrics can also be applied against the generated statistics to create key performance indicators or to establish costs or weights to errors. You can track and trend any of those measures over time.

Baseline analysis—Comparing results from current profiling sessions with the results from a prior profiling session offers insight into what may have changed in structure or content over time.

Evaluation of source-to-target transformation—InfoSphere Information Analyzer Exception Manager can continuously assess the underlying transformation rules automatically found via IBM InfoSphere Discovery or any rule that can be manually described in SQL and identify exception conditions for remediation.

Data steward interface for remediation—InfoSphere Information Analyzer Exception Manager enables you to manage the exception life cycle: monitor rule exceptions, propose changes to exception data, ignore repeat exceptions and ensure that data quality conditions are maintained.

Extensive reporting—More than 80 reports are ready to use out of the box. Visual representation of analysis, trends and metrics and incorporation of user notes all help you understand results quickly and efficiently. Newly introduced application programming interfaces (APIs) and command-line interfaces (CLIs) provide the ability to integrate results in your own custom environments.

Ability to scale up to handle high volumes of data—Native parallel data profiling and rules analysis enable you to easily analyze and evaluate high volumes of data.

Secured analysis—Project-, role- and user-based security controls protect access to sensitive analytical information. Controls can be as open or as restrictive as needed to meet compliance requirements.

Collaborative multi-user environment—With support for multiple analytical reviews as well as asynchronous profiling, a project-based environment allows more than one user to work in a project-based context.

Integrated metadata—Analytical results and notes can be automatically shared across all InfoSphere Information Server modules. Integration with InfoSphere Business Glossary helps you connect data sources to your business terminology so that insights into data quality can be provided to a broader audience.

Broad support for heterogeneous data sources—InfoSphere Information Analyzer can profile IBM DB2®, Oracle, Microsoft® SQL Server, Sybase, IBM Informix®, Microsoft Access and other data sources (such as text files) through Open Database Connectivity (ODBC), or utilize native connectivity to DB2, Oracle and Teradata databases.

With InfoSphere Information Analyzer, you can leverage an understanding of source systems to uncover anomalies including:

- Missing data
- Inaccurate and inconsistent data
- Duplicate data
- Exceptions to data rules
**Data profiling within a unified platform increases accessibility and consistency**

Leveraging a Service Oriented Architecture (SOA) to unlock information from individual silos, InfoSphere Information Analyzer delivers data profiling and rules analysis functions within the context of a complete information integration platform, enabling greater accessibility and consistency throughout the enterprise.

Active metadata across InfoSphere Information Server helps simplify the collection and management of metadata across the entire integration spectrum. Profiling, analysis and monitoring results within InfoSphere Information Analyzer are stored in the common metadata repository. The result? Data quality information pertinent to users of IBM InfoSphere DataStage® and InfoSphere QualityStage software is instantly accessible without the need to export or import the data.

**InfoSphere Information Server delivers trusted information**

Organizations face tremendous challenges in locating information, finding it in the form needed and discerning insights once it is found. Information validity and control are additional concerns. The challenges only mount if businesses cannot ensure access to authoritative, consistent, timely and complete information.

InfoSphere Information Server is a revolutionary 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 source. It provides breakthrough productivity and performance for cleansing, transforming and moving this information consistently and securely throughout the enterprise, so that it can be accessed and used in new ways to help drive innovation, help increase operational efficiency and help lower risk.

InfoSphere Foundation Tools—including InfoSphere Information Analyzer—help you put your arms around your data by enabling companies to understand disparate data spread across their heterogeneous systems, govern it as business information over time and design trusted information structures for business optimization. InfoSphere Foundation Tools work with almost any data integration, business intelligence or data warehouse tools. They also work in conjunction with the comprehensive set of solutions for complete end-to-end data integration.

**System requirements**

InfoSphere Information Analyzer supports IBM AIX®, Red Hat Enterprise Linux®, Novell SUSE Linux Enterprise Server, Solaris, HP-UX and Microsoft Windows® operating systems. InfoSphere Information Analyzer Exception Manager is supported on Microsoft Windows operating systems. For current, detailed hardware and software system requirements for IBM information integration products, visit: [ibm.com/software/data/infosphere/information-analyzer/requirements.html](http://ibm.com/software/data/infosphere/information-analyzer/requirements.html)
For more information
To learn more about IBM InfoSphere Information Server and IBM InfoSphere Foundation Tools, please contact your IBM marketing representative or IBM Business Partner, or visit: ibm.com/software/data/ips