IBM Information Management software

What's new in IBM InfoSphere Information Server 8.1

Contents

Introduction

IBM InfoSphere™ Information Server® is a revolutionary software platform that helps organizations derive more value from the complex, heterogeneous information spread across their systems. It provides breakthrough collaboration, productivity, and performance for cleansing, transforming, and moving this information consistently and securely throughout the enterprise. It can then be accessed and used in new ways to drive innovation, increase operational efficiency, and lower risk.

Figure 1: InfoSphere Information Server capabilities overview
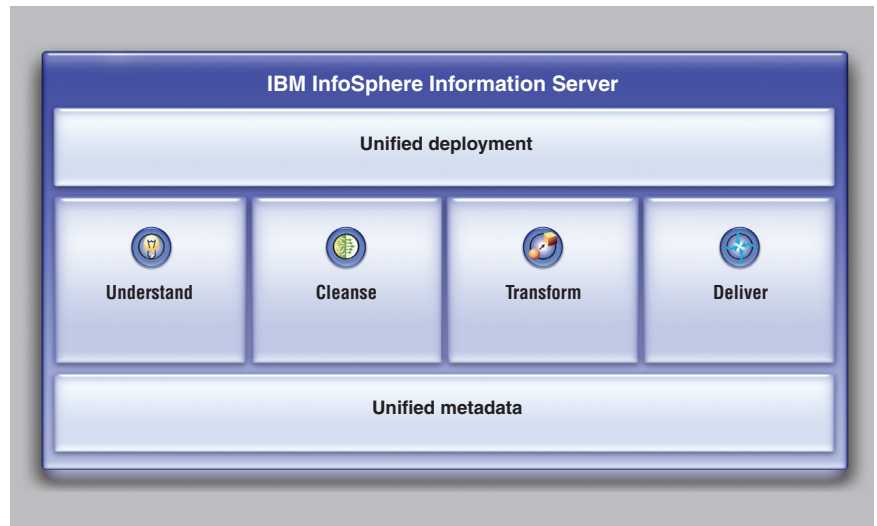
This paper outlines the new enhancements and features in the latest release of Information Server 8.1. For additional information, please reference the individual What's New documents for the following Information Server products:

- *What's New in IBM InfoSphere DataStage® and QualityStage® 8.1*
- *What's New in IBM InfoSphere Business Glossary® and Metadata Workbench® 8.1*

Summary

IBM Information Server combines a variety of market leading technologies into a single unified platform that enables companies to understand, cleanse, transform, and deliver trustworthy and context-rich information. Information Server 8.1 includes enhancements and new features to the following core capabilities (see Figure 1)

- *Understand*
    - o   *IBM InfoSphere Business Glossary*
    - o   *IBM InfoSphere Business Glossary Anywhere*
    - o   *IBM InfoSphere FastTrack*
    - o   *IBM InfoSphere Information Analyzer*
- *Cleanse*
    - o   *IBM InfoSphere QualityStage*
- *Transform*
    - o   *IBM InfoSphere DataStage*
- *Deliver*
    - o   *IBM InfoSphere Information Services Director*

In addition, significant enhancements have been made to the unified platform, connectivity, and IBM InfoSphere Metadata Workbench.

Helping customers understand their information

Business Glossary

Business Glossary enables business analysts and subject-matter experts to create, manage, and share a common enterprise vocabulary and classification system. This enables users to link business terms to more technical artifacts managed by Information Server. These business terms are shared with IBM InfoSphere Data Architect (previously named Rational Data Architect), Information Analyzer, and FastTrack creating a common set of semantic tags for reuse by data modelers, data analysts, business analysts, and end users.

What's new in Business Glossary 8.1

- *Enhanced naming and category tree scalability for large business glossaries*

Because organizations often have multiple definitions for the same category or term within different departments, Business Glossary now supports having multiple terms and categories with the same name under different parent categories. This ensures you can accurately classify organizational information into hierarchies which represent complex business relationships. In addition, Business Glossary delivers improved support for large business glossaries that have many categories and terms which represent complex hierarchies such as the IBM Industry Models. This allows organizations to start small with implementations and ensure support for expanded implementations as their glossary usage spreads across the enterprise to thousands of users.

- *New control support for term access*

Terms can be assigned a status of standard, accepted, candidate and deprecated within their lifecycle. Administrators can optionally restrict access to these terms by status and control their visibility within Business Glossary Browser and Business Glossary Anywhere interfaces. For example, organizations may opt to restrict displaying any candidate terms until they are approved for wider usage.

- *Improved user productivity*

Enhanced batch editing capabilities allows changes to be applied to multiple objects at one time rather than individually. This enhancement drastically increases productivity of maintaining Business Glossary contents by

shortening the time to apply updates to content and reduces the chance of errors when manually applying changes to many objects at one time. Multiple enhancements have been made to streamline searching for metadata assets and classifying linkages between business and technical metadata. Classifying objects through a batch process versus manually reduces the chance of errors and improves the accuracy of the glossary's contents.

Enhanced search ranking algorithms allow candidate results to be listed in best-matched order instead of straight-alphabetical, helping business users find the right match faster and easier.

- *Simplified import*

Multiple enhancements have been made to expand and simplify the CSV import capabilities for creating and managing the business glossary contents. By allowing Business Glossary administrators to easier leverage existing fragmented, home grown documentation and glossaries, the breadth of information supported by a single glossary can reach a much wider business audience.

Figure 2: InfoSphere Business Glossary Simplified import
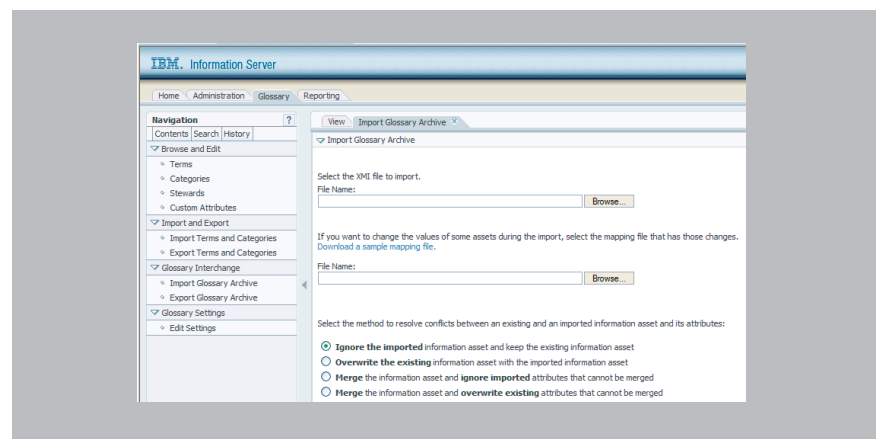
• *Expanded support for external application URLs*

Valuable metadata is often stored in external documents and application but still needs to be referenceable and accessible to provide a complete picture. The ability to include links to external sources such as web pages, word documents, local file servers, etc. is now supported. These links can also be included in any text field and are clickable and will open in a separate window displaying the linked content. By incorporating the ability to link to external documents, organizations can expand on data governance initiatives and link Information Server assets to additional sources and enterprise documents.   This capability is also available in Business Glossary Anywhere.

Business Glossary Anywhere

Business Glossary Anywhere enables users to instantly understand business terminology. Directly accessible from any desktop application, users click on a term and view its business definition in a pop-up window without loss of context or focus. This simple, direct access mechanism to Business Glossary definitions promotes user productivity and consistent application of business terminology across organizations.

What's new in Business Glossary Anywhere 8.1

• *New enhanced search capabilities*

Business users can now quickly search for category information as well as business terms with the click of a button. This access to additional information promotes ease of use and increased adoption of a standard business glossary as part of the business user's workflow process.

• *New navigation linkage to support summary and detailed information*

User can now go directly from Business Glossary Anywhere to the equivalent Business Glossary browser page for additional information on business categories or terms searched. Ease-of-access to more detailed information about business glossary promotes expanded usage and streamlines navigation of contents.

- *Enhanced application deployment*

Deploying Business Glossary Anywhere for thousands of enterprise users is made easier with the option to place the installable executable on a shared file server to update remote machines automatically the next time a client logs into their Business Glossary Anywhere application. This simplified deployment mechanisms makes managing thousands of users virtually maintenance free.

FastTrack

FastTrack accelerates the deployment of data integration projects by focusing on what historically has been a manual translation of business requirements into a technical implementation. It provides a managed infrastructure and spreadsheet-style interface for the business analyst to create source-to-target specifications and links transformation rules to source data, target models and business terms. It generates Data Stage jobs with in-line, instruction-specific annotations. FastTrack promotes collaboration between business and technical users, improving overall efficiency and performance.

What's new in FastTrack 8.1
- *New Application of Corporate Naming Standards*

Naming standards can be created and defined for each FastTrack project. These standards are automatically applied during the generation of Data Stage jobs and can be applied at both the stage and link level. In addition, Business Glossary standard abbreviations can also be the basis for these naming standards. The ability to apply common naming standards allows organizations to enforce best practices, consistency and readability across documentation and Data Stage jobs.

Figure 3: InfoSphere FastTrack new application of corporate

naming standards
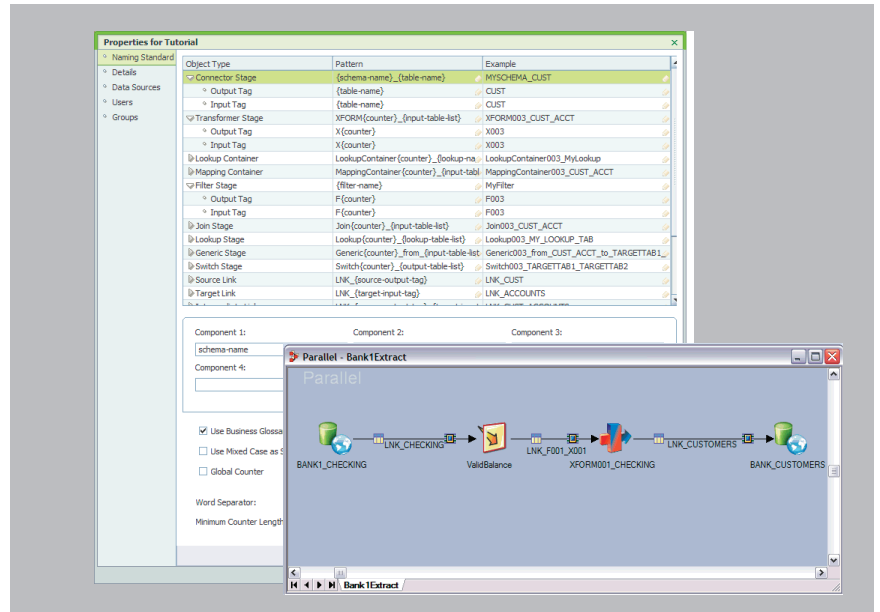
• *New substitution of data model structures from development to deployment*

As specifications are completed during development and are ready to migrate to new environments such as QA/testing or production, data model structures can be switched to use new host, database, schema, table and column names using a matching replacement wizard. This process significantly minimizes manual work required to migrate specifications and Data Stage jobs between multiple environments and reduces chance of errors being introduced.

Figure 4: InfoSphere FastTrack new substitution of data

model structures
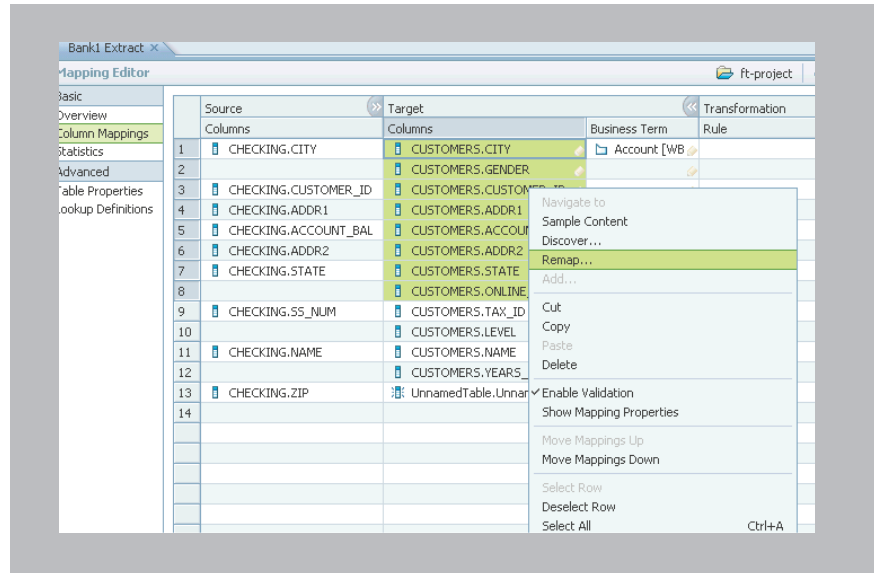
- *Expanded support for database and file connector options*

When defining a FastTrack specification connectivity parameters, users can now select from expanded database connectors including DB2®, Flat Files, Oracle and Teradata. This capability provides organizations with the flexibility to select optimal source and target connectors for best performance.

- *Enhanced specification reporting*

FastTrack specification reporting is available from either the FastTrack client or from the Information Server web console and includes support for Japanese translation. This capability ensures superior language support for organizations with global business requirements and promotes accessibility of documentation.

- *Improved integration with the Metadata Workbench*

FastTrack mapping specifications can now be viewed within the Metadata Workbench, providing tighter integration between the products and delivering stronger unified metadata support for our customers. The ability to view FastTrack metadata assets in the Metadata Workbench provides increased productivity and more cross visibility of shared metadata to our customers.

Information Analyzer

Information Server brings a next-generation data profiling and analysis tool to market with its Information Analyzer product that is designed to help business and data analysts understand the content, quality, and structure of their data sources by automating the data discovery process.
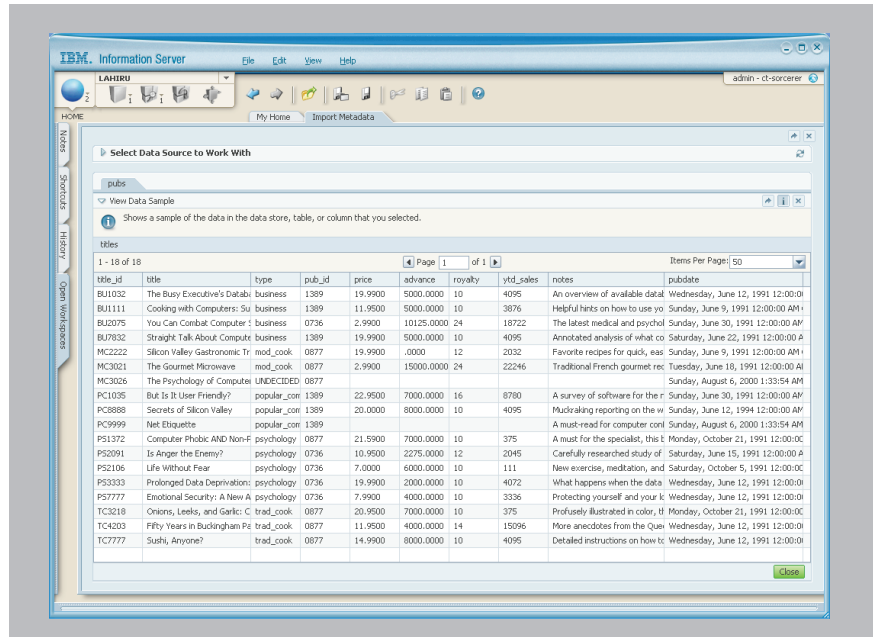
What's new in Information Analyzer 8.1

- *Enhanced Understanding & Usability*

The profiling functionality in Information Analyzer 8.1 provides easier and more fine-grained understanding of data. Paging and sorting enhancements as well as new filtering options make it easy to review very large sets of profiling results as well as provide focus on specific data segments of interest to an analyst. A greatly improved data drilldown allows the analyst to easily shift between the analysis detail and the drilldown information.

This release also allows the analyst to view business metadata directly in Information Analyzer so the analyst does not need to rely on the technical data definitions. A new Metadata Import Preview and Filtering feature allows the user to have a quick view into the data content before importing metadata to ensure that the import targets the right data source.

Figure 5: InfoSphere Information Analyzer enhanced understanding and usability


- *Global enablement and extended quality analysis*

With the Information Analyzer 8.1 release, full support in all screens and reports is available for double-byte character sets allowing customers to profile and view the information in their global data sources. Additional enhancements allow for recognition of locale-specific dates and number formats as well as presentation in the user interface based on the specific locale.

Further, this release also introduces new format-driven data validation that allows the analyst to quickly identify incomplete or invalid values based on non-conforming formats. The analyst can address quality considerations faster for their enterprise data.

- *Focused reporting*

Information Analyzer 8.1 introduces context-sensitive reporting; allowing immediate access to reports from within function screens, and adds many new analytical reports to its wide range of existing reports. All analytical information can now be readily generated and delivered to users in standard formats through the user interface or the Information Server reporting console.

Figure 6: InfoSphere Information Analyzer focused reporting
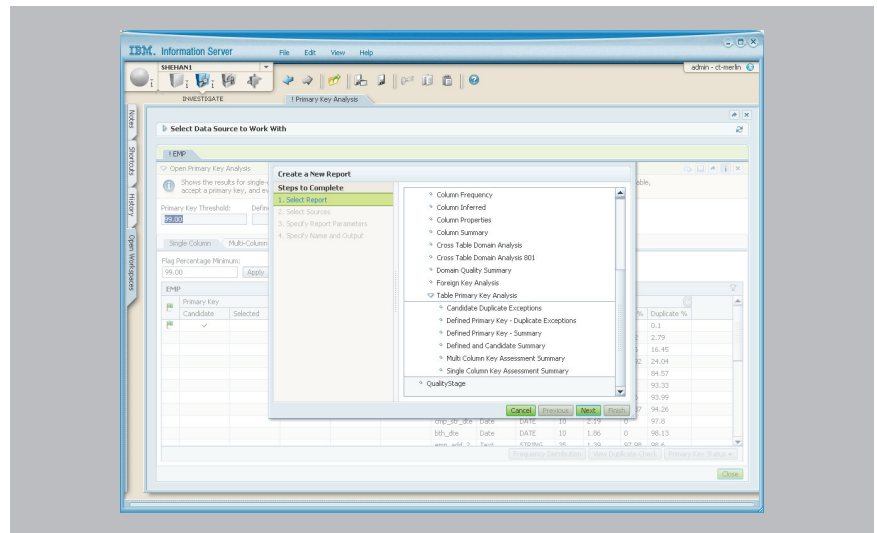
- *Extended execution control*

Finally, this release continues to provide execution enhancements through simultaneous multi-job processing that leverages the native parallel architecture; source selection refinements in both foreign key and cross-domain analysis; and a free-form where clause option for column analysis. These changes allow the analyst greater control in profiling execution, focusing only on those data segments of interest.

Next generation of data quality and transformation capabilities

DataStage and QualityStage
Information Server brings next-generation data transformation and data quality capabilities to the market with its DataStage and QualityStage products. DataStage enables organizations to design data flows that extract information from multiple source systems, transform it in ways that make it more valuable, and then deliver it to one or more target databases or applications.   QualityStage provides core data quality functionality, delivered as part of Information Server, to help improve the quality of the organization's information.

What's new in DataStage and QualityStage 8.1

• *Enhanced deployment*

The Information Server Manager has new deployment functionality that helps customers move artifacts between environments in an auditable and secure way. In prior releases (DataStage 7.5 and before), DataStage Version Control was used to "batch promote" jobs between environments. The new Manager goes beyond what DataStage Version Control had, to provide a simpler, auditable and secure mechanism to promote between environments.

Figure 7: InfoSphere DataStage and QualityStage enhanced deployment
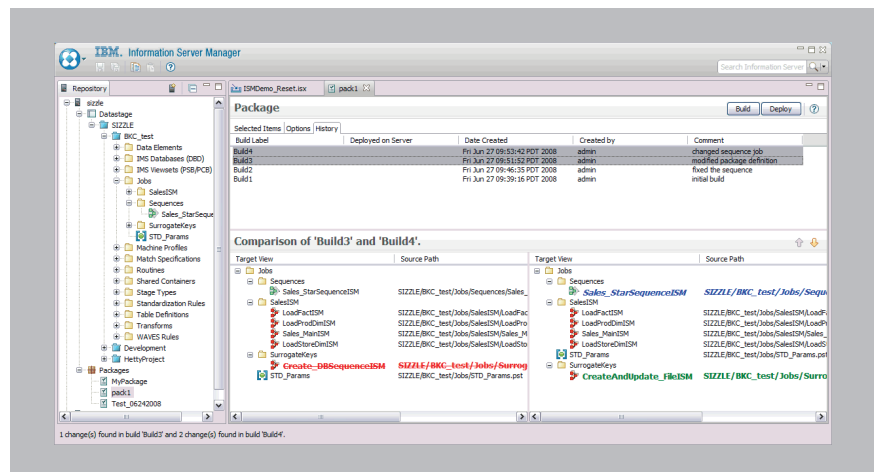
The graphical user interface lets users connect to multiple projects in any Information Server environment. A description of a "package" can be created with any first class objects from the repository. Other associated objects can be added in such as files outside of DataStage and QualityStage, like scripts. New objects can be added to an existing package. This description is stored in the metadata repository.

Users can then build the deployment package through the user interface or using the command-line interface. The command-line interface can now also be run on the Windows client or the server. Once the packages are created, they can be deployed from the Manager or command line. The Read-only objects and executables (including transformer executables, buildops, etc.) can be moved too so it's not necessary to compile on the production systems.

DataStage and QualityStage users with access to the projects can create packages. Only users with a Production Manager role can deploy to the target system. This is done to ensure only users with the right privileges can modify the target (production) environment.

The Manager also maintains a record of all the deployments made, including who performed the action, when it was done, and all the objects that were moved in that deployment. This provides users with the ability to track any movement into production systems.

The Information Server Manager also gives you the ability to roll back to a previously deployed package. The user can choose a previous deployment compare the artifacts in that deployment with the latest one deployed and choose to go back to the previous version.

The Information Server Manager has the ability to compare any two deployments. It shows details of new objects, objects that were not in the next build and objects that changed between the two deployments.

- *Column-level impact analysis*

Column-level impact analysis is available in the 8.1 release of DataStage and QualityStage. This capability allows users to search for occurrences of column names within Table Definitions, Jobs or Shared Containers, as well as perform a "where used" analysis on a column and see the jobs, stages and links where a selected column is used. This new capability brings enormous productivity benefits to DataStage and QualityStage users as it allows for an easier and faster assessment of impact analysis.

- *Logging View from web browser*

A new web-based Logging View allows users to view DataStage and QualityStage production logs from Information Server's web-based Administration Console without having to have access to the production systems. Users can choose from fully customizable templates to create a user defined view. With a new user role for the logging view, specific users, such as developers, can debug logs on secure production systems without needing access to DataStage and QualityStage Director or Designer.

- *New Pivot and Check Sum Stages*

The new Pivot Stage in DataStage supports horizontal pivoting and splits each input row into multiple output rows, repeating the non-pivoted data in each output row.

The new Checksum Stage allows error checking to ensure the accuracy of transmitting digital data by computing a checksum of a set of input fields and placing this checksum in the output along with the original fields. Both stages support parallel execution delivering high performance for these operations.

- *DataStage Balanced Optimization*

A new option for DataStage Customers, DataStage Balanced Optimization provides the ability to push transformation into the source database(s), process in DataStage's high performance engine, and/or push processing into the target database(s). Users will be able to continue to express the logic of their integration processes in current natural flow-oriented DataStage convention, and then automatically optimize their designs to achieve the benefits of pushing some processing to the source and/or target databases. This new capability will not require users to manually rewrite their DataStage jobs to achieve these benefits. This capability will allow users to choose the benefits of "ETL", "TELT", and even "TETLT" without sacrificing the benefits of DataStage.   For more information, please reference the IBM whitepaper on DataStage Balanced Optimization.

Figure 8: InfoSphere DataStage Balanced Optimization
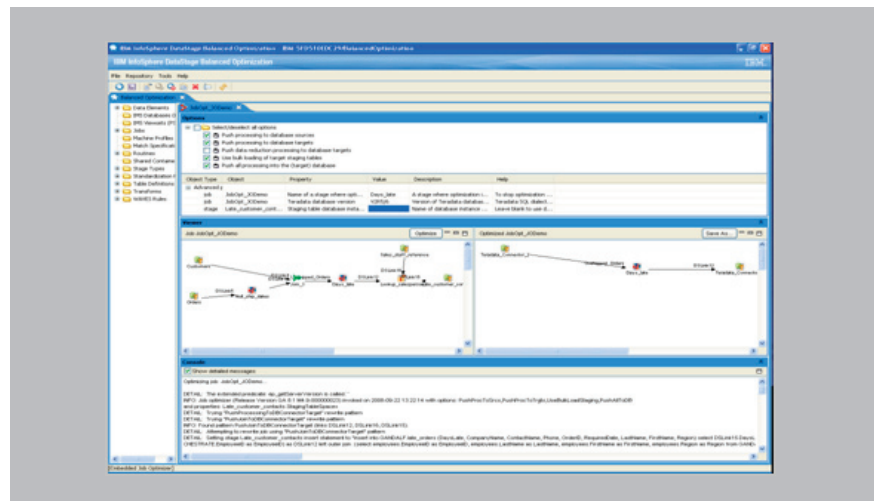
- *Address verification and Match Enhancements in QualityStage*

QualityStage 8.1 expands the data quality rule sets with address verification for 240 countries. QualityStage address verification can standardize, verify, and correct multinational address data against postal reference files, for better information integration as well as customer and partner communications worldwide.

- *Match Enhancements in QualityStage*

QualityStage 8.1 delivers a new Match Type – Unduplicate Transitive Match. The Transitive option will appear in the Unduplicate Match Stage Editor and in the Match Designer GUI. This new Match Type builds groups within a block using transitive matching. So if A matches B and B matches C, then A, B, and C are all in one group even if A doesn't match C. This Match Type extends the transitive grouping of Undup Independent to inside of a pass. This capability extends QualityStages versatility by enabling customers to allow records to match indirectly within a pass, helping to overcome the ever present challenges of poor quality data.

QualityStage 8.1 also delivers a new Match Comparison Type – MULT_ALIGN. The MULT_ALIGN comparison scores the similarity of two sequences of terms. This comparison combines your knowledge of how similar the terms are, the order of the similar terms, and the proximity of the similar terms. This capability extends QualityStage's ability to deliver superior match results for complex multi value domains.

QualityStage 8.1 also delivers new capability to define Logical Group Definitions within a Match Specification. A group is composed of a set of match fields which is defined by specifying a list of fields or already defined groups. Group definitions allow a QualityStage user to define match rules to support more complex business logic in a more simplified manner buy enabling the specification of special handling to occur at a group level rather than field. This capability enables more productive development, control and maintenance of match rules.

Delivering trusted information

Information Services Director

Information Services Director provides a unified mechanism for publishing and managing shared services across data quality, data transformation, and federation functions, allowing information specialists to easily deploy services for any information integration task and consistently manage them. This enables developers to take data integration logic built using Information Server and publish it in minutes as an "always on" service.

What's new in Information Services Director 8.1

Figure 9: InfoSphere Information Services Director bindings and information providers
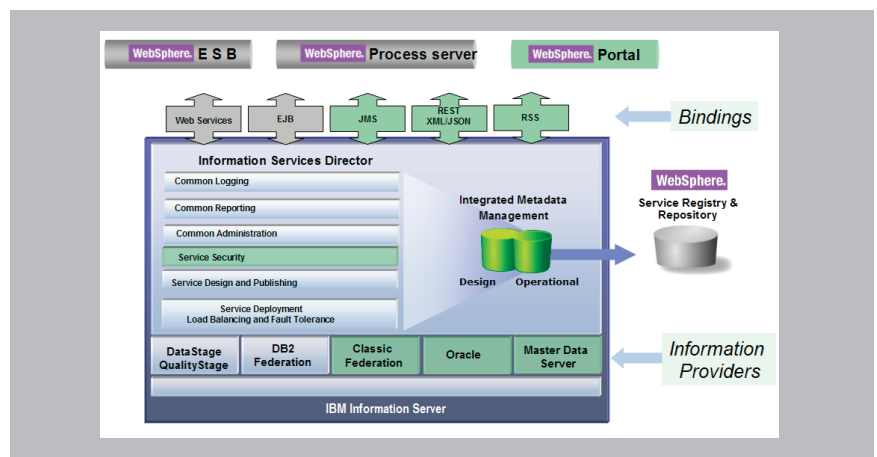
- *New bindings*

New bindings in the Information Services Director include JMS (support for SOAP or Text over JMS), RSS (support for delivery of RSS formatted data as services), and Representational State Transfer (REST support for XML and JSON formatted data over REST). These new bindings enhance the utility, versatility and reuse of services, especially in a Web 2.0 environment, and allow organizations to deploy information as a service in a way that is most comfortable and that utilizes their investment in other service-oriented architecture infrastructure (SOA) components.

- *New information providers*

This release supports service publishing from a variety of new sources, including Oracle databases, InfoSphere MDM Server and mainframe data sources like VSAM, Adabas, IMS, and IDMS through InfoSphere Classic Federation Server. This capability provides broader utilization of information assets for an organization while extending the capability of Information Server as the centralized repository for all information as a service activities.

- *Improved security features*

This release provides the ability to enable service authentication and confidentiality delivering stronger security capabilities to ensure only appropriate and authorized users can access and use specific information sources.

- *Integration with Master Data Management Server*

Services created and deployed on Master Data Management Server can be registered with Information Services Director. This capability centralizes control and distribution of information services from a single platform for greater data governance. This integration also allows master data management services from Master Data Management Server to be utilized and integrated with other SOA foundational products such as IBM WebSphere Process Server, IBM WebSphere Portal Server, and IBM WebSphere Service Registry and Repository via the direct integration capabilities of IBM Information Services Director.

Unified Metadata Management

Metadata Workbench

Metadata Workbench offers key metadata visualization and exploration capabilities for Information Server. Acting as a control station, it provides cross-tool impact analysis and data lineage, extensive reporting and metadata administration functions for Information Server.

What's new in Metadata Workbench 8.1

- *New data lineage viewable from within Cognos Framework Manager 8.4*

View Information Server data lineage directly from within Cognos Framework Manager 8.4 without logging separately into the Metadata Workbench browser. This new capability allows business users to immediately access and understand data lineage from within Cognos reports to increase confidence in reporting decisions.

- *Enhancements to data lineage and impact analysis reporting*

A variety of improvements have been made to viewing and reporting on data lineage and impact analysis within the Metadata Workbench. When invoking analysis reports, users may select the type (operational, design, or user-defined) of metadata path to be viewed. Also, visual enhancements have been made to demonstrate the flows of these different types of metadata. Reporting capabilities have been improved to show all related metadata assets and this information may be saved to a file or printed.

Figure 10: InfoSphere Metadata Workbench enhancements

to data lineage and impact analysis
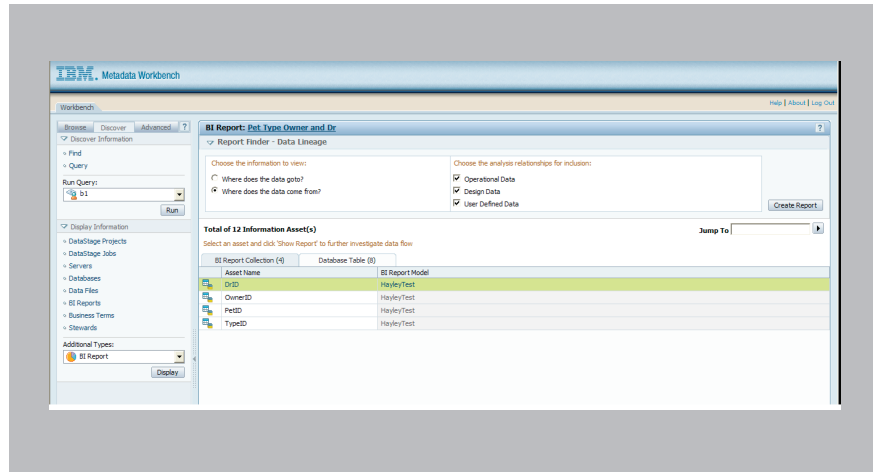
- *New full query lifecycle support*

The query builder feature allows users to create, export, save and publish queries on metadata stored in the Information Server repository. Query results can be exported in HTML, .xls, and .csv format. This feature improves metadata management and information access across all installed Information Server components. Reusable queries support change management, data governance initiatives, and monitoring of key areas of the metadata.

Figure 11: InfoSphere Metadata Workbench full query

lifecycle support
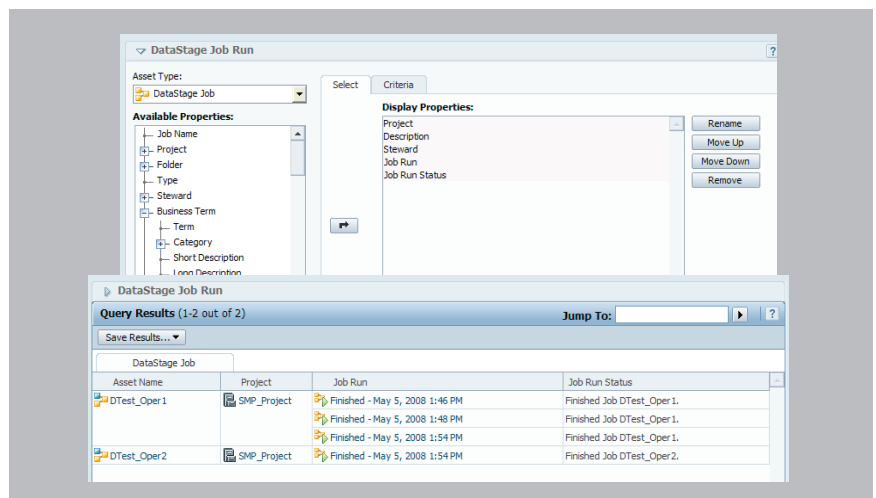
- *New user roles for enhanced security*

Multiple new user roles were added to provide additional control over the Information Server platform metadata management and access levels. The two new roles are the Metadata Workbench User and the Metadata Workbench Administrator. This finer-grained security allows organizations the flexibility to customize management of metadata for their specific business requirements.

Platform and connectivity enhancements

Platform
Information Server is the industry's first comprehensive, unified platform for enterprise information architectures, capable of scaling to meet any information volume requirement so that companies can deliver business results within these initiatives faster and with higher quality results.

New platform capabilities in Information Server 8.1

- *Globalization*

Information Server 8.1 will allow users to leverage this revolutionary software platform in their native language. The console and user interfaces, messages, and most documentation will be translated into 9 languages for DataStage, QualityStage, Information Services Director, Business Glossary, and the Metadata Workbench. Language support includes: simplified Chinese, traditional Chinese, Japanese, Korean, French, German, Italian, Spanish, and Brazilian Portuguese.

- *Native 64-bit Support*

Information Server 8.1 continues to add native 64-bit support to its wide range of supported operating systems. This release provides native 64-bit support for AIX and Solaris in addition to the native 64-bit support for HP-UX on Itanium and Linux on System Z, which was delivered in 8.0.1. Information Server 8.1 (and 8.0.1) also supports 64-bit Red Hat Enterprise Linux, SUSE Linux Enterprise Server and HP-UX on PA-RISC, but remains a 32-bit application on these platforms. For a complete list of platforms supports by Information Server, please reference the announcement letter at www.ibm.com.

Connectivity

Information Server enables tight integration of enterprise information, regardless of the sources, targets and time frames. Whether for building an enterprise data warehouse, or a "real-time" operational reporting data mart, or integrating dozens of disparate source systems to support strategic business systems like CRM, SCM, ERP, BI/analytics, and e-business, Information Server helps ensure the success of your enterprise data integration initiatives via optimized connectivity.

Information Server delivers all the capabilities necessary for success in enterprise data integration: the ability to quickly and easily connect to the most comprehensive list of supported sources and targets, advanced maintenance and development, which simplifies administration and speeds implementation, and the most scalable platform available, to handle today's massive volumes of new corporate data through high-performance processing and data movement.

New connectivity capabilities in Information Server 8.1

- *Improved connectivity for databases and files*

Included in Information Server 8.1 is a new DB2/Z Stage, this provides DataStage & QualityStage with native, bulk load and unload capabilities for DB2 on System Z.

Additionally, the integration between Classic Federation and Classic Event Publisher offerings and DataStage has been further optimized with more direct exchange of data (no longer requiring WebSphere MQ Series) via a raw data format, delivering significant performance improvements. Other new database connectivity enhancements include support for:

- *IBM DB2 v9.5*
- *IBM IDS 11*
- *IBM WebSphere MQ Series v6.1*
- *Teradata v12 and 13*
- *Oracle 11g*
- *SQL Server 2008*
- *Netezza 4.0 and 4.5*
- *MySQL v5.0+*

These database connectivity enhancements are delivered with Information Server 8.1 or earlier.

- *Optimized integration with Changed Data Capture (CDC)*

Information Server 8.1 also includes optimized integration with IBM InfoSphere Change Data Capture (formerly known as DataMirror), which enables very low impact, log-based sourcing (from DB2 on all platforms, Oracle, Sybase and SQL Server) with direct, real-time feeds into DataStage jobs.

Informix and Sybase log-based capture is planned for the next version of CDC. By providing the ability to process (source/transform/apply) only changed data, CDC enables Information Server to address the exponential explosion of enterprise data and shrinking batch windows, while also extending Information Server's real-time, transactional capabilities with event-driven processing. This future release of CDC will also extend its language support to include Japanese.

- *Connectivity enhancements to enterprise applications*

All previously released Packs for SAP R/3, SAP BW, Oracle Applications, Siebel, PeopleSoft and JD Edwards have been updated for globalization and native 64-bit support. A new Pack for Salesforce.com (v1.0.1) is available that provides a certified, metadata-based, fully configurable, high performance data integration between salesforce.com and other enterprise data repositories.

A new MapStage to support the latest release of WebSphere Transformation Extender has also been delivered, providing access and mapping for more complex, semi-structured data including EDI.

Finally, the Pack for Oracle Applications now supports NoetixViews (developed by Noetix Corp.). This combination enables highly intuitive connectivity to Oracle's E-Business Suite, leveraging simplified views of underlying application schema documented in meaningful business terms to the Information Server job developer.

Migration
IBM is committed to helping customers migrate to new products versions in the most seamless and cost effective way possible. Customers use IBM products to run their business, so fast and low cost migrations are important to ensuring the regular delivery of new and improved technology.

Additional information

For more information about IBM Information Server, contact your IBM marketing representative or visit **ibm.com**/software/data/integration

Also for additional information about what's new in Information Server 8.1, please reference the following documents:

- *What's New in IBM InfoSphere DataStage and QualityStage*
- *What's New in IBM InfoSphere Business Glossary and Metadata Workbench*

IBM®

TAKE BACK CONTROL WITH Information Management