IBM Information Server, Version 8.0: Transforming and delivering data

---

Highlights

■ **Build a scalable infrastructure to support a single view of enterprise information**

■ **Easily access and transform data from multiple source systems, and deliver trusted information in a consistent and consolidated form—wherever and whenever it is needed**

■ **Achieve breakthrough productivity by eliminating hand-coded data integration programs**

■ **Improve information quality by applying the same transformation rules for batch, real-time or on demand integration services**

■ **Reduce the costs and risks associated with complex IT initiatives**

Create a single view of enterprise information

Every day, torrents of data inundate IT organizations and overwhelm the business managers who must sift through it all to glean insights that help them grow revenues and optimize profits. Yet, after pouring hundreds of millions of dollars into new enterprise resource planning (ERP), customer relationship management (CRM), supply chain management (SCM), business intelligence (BI), business process management and data warehousing systems, many companies are still plagued with disconnected, "dysfunctional" data—a massive, expensive sprawl of disparate silos and unconnected, redundant systems that fail to deliver the desired single view of the business.

As a first step in deploying new enterprise applications, integrating a new acquisition or tackling an IT business process initiative, it is absolutely critical to ensure that the underlying data is reliable, relevant and readily available in a consistent, consolidated form—wherever and whenever it is needed—across the enterprise. IBM® Information Server addresses the critical need for data transformation, enabling organizations to design an infrastructure for data integration that is scalable, reliable and flexible for meeting the challenges of today's dynamic business environments.

IBM Information Server enables organizations to tightly integrate enterprise information, despite having many sources and targets, large data volumes and short timeframes.

ON DEMAND BUSINESS™

Whether building an enterprise data warehouse to support the information needs of the entire company, building a real-time data warehouse or integrating dozens of source systems to support enterprise applications like CRM, SCM and ERP, IBM Information Server helps ensure that organizations will have information they can trust (see Figure 1).

Key capabilities of IBM Information Server

IBM Information Server delivers three key capabilities necessary for success in information integration: comprehensive connectivity to easily and quickly access any source or target system; advanced development and maintenance tools, which speed implementation and simplify administration; and a scalable platform that can easily handle today's massive volume of corporate data.

IBM Information Server supports the collection, integration and transformation of data structures ranging from simple to highly complex. It easily manages data requiring attention within seconds, as well as massive quantities of data that need daily, weekly or monthly processing intervals.
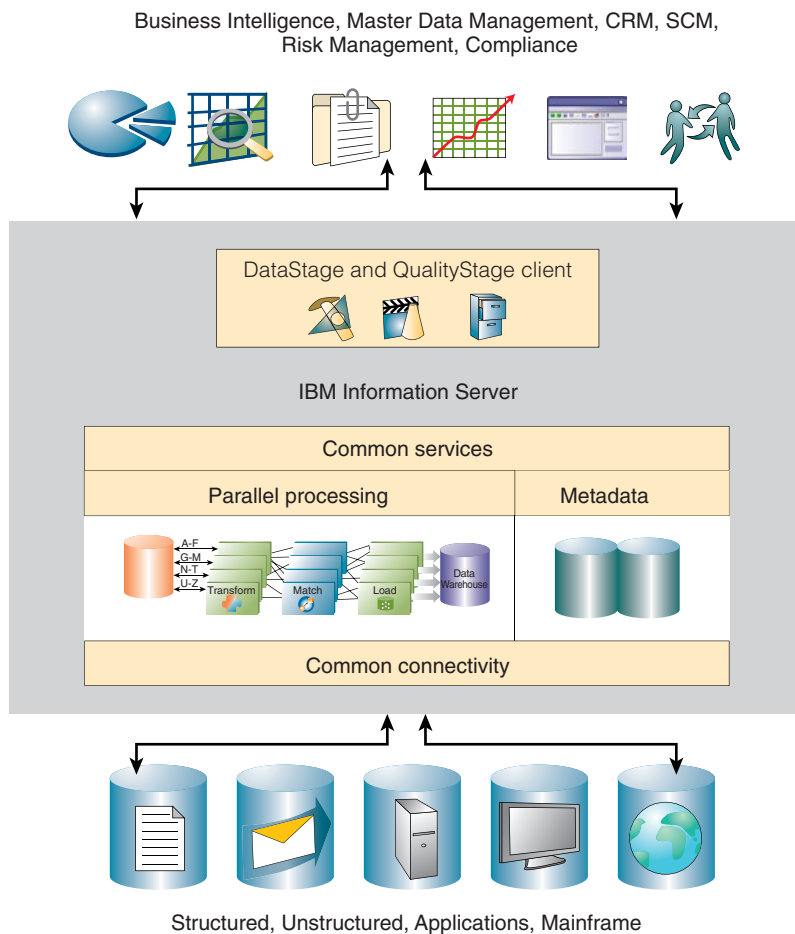
Enterprise connectivity

Enterprise-class information integration requires access to all data sources within the enterprise. IBM Information Server enables a virtually unlimited number of heterogeneous data sources and targets to be connected and combined within a single job.

Business Intelligence, Master Data Management, CRM, SCM, Risk Management, Compliance

Structured, Unstructured, Applications, Mainframe

Figure 1: IBM WebSphere DataStage enables access to multiple source systems, integrating and transforming selected data to deliver trusted information to critical business functions.

2

These sources and targets may include:

- *Text files including complex mainframe files*
- *Complex XML data structures*
- *Enterprise application systems such as SAP, Siebel, Oracle and PeopleSoft*
- *Almost any database, including partitioned databases such as Oracle, IBM DB2 Universal Database™ (with and without Data Partitioning Feature), IBM Informix®, Sybase, Teradata and Microsoft® SQL Server™*
- *Web services*
- *SAS*
- *Messaging and enterprise application integration (EAI) products including IBM WebSphere® MQ*

Advanced development and maintenance

IBM Information Server utilizes a powerful architecture that gives developers maximum speed, flexibility and effectiveness in building, deploying, updating and managing their data integration infrastructure.

The productivity-enhancing features delivered with IBM Information Server help reduce the learning curve, simplify administration and optimize the use of development resources, resulting in a decreased development and maintenance cycle for data integration applications. As a result, IBM Information Server enables companies to spend less time developing integration processes and more time reaping the benefits of having trusted information shared across their applications and databases.

Data integration and data quality

IBM Information Server has a single design interface that is shared by Data Transformation and Data Quality product modules. Because of this interface, designers can use any combination of the data quality and data transformation capabilities to help ensure that the right data is brought together at the right time. Since both product modules also share a unified metadata repository, the critical business, technical and process metadata that is developed during data profiling, data quality or data integration processes is immediately available from any of the product modules. This design greatly speeds development and reduces the chance for errors.

Ease of use

The unified design environment employs a work-as-you-think design metaphor. Developers use a top-down data-flow model of application programming and execution, which allows them to create a visual data flow (see Figure 2). A robust graphical palette helps developers diagram the flow of data through their environments via simple, GUI-driven, drag-and-drop design components. More than 50 prebuilt components and hundreds of transformations are included for maximum productivity. Developers also benefit from powerful debugging capabilities and an open application programming interface (API) for leveraging external code.

Productivity and reuse

IBM Information Server shortens the development cycle by promoting the reuse of existing data integration business logic. This process employs the concept of containers in which jobs and metadata created in one container can be shared and reused by other jobs. Quick and Advanced Find capabilities make it easy to locate objects for reuse across different projects. Robust job specification reporting provides documentation so other developers can easily understand job design and provide additional support.

Right-time data integration

IBM Information Server can operate in real time, capturing messages or extracting data at a moment's notice on the same platform that also integrates bulk data, using the same transformation rules. In addition, data integration jobs can easily be deployed as Java™ Message Services, JavaBeans™ or as Web services. This approach provides a key advantage by enabling complex data integration processes to be shared with a much wider set of developers who do not need to understand the complex steps contained in the services. The result is that data can be used in more ways—without costly hand coding—to respond to an organization's information integration needs on demand. Other offerings may require the use of two or more separate tools to achieve the same functionality.
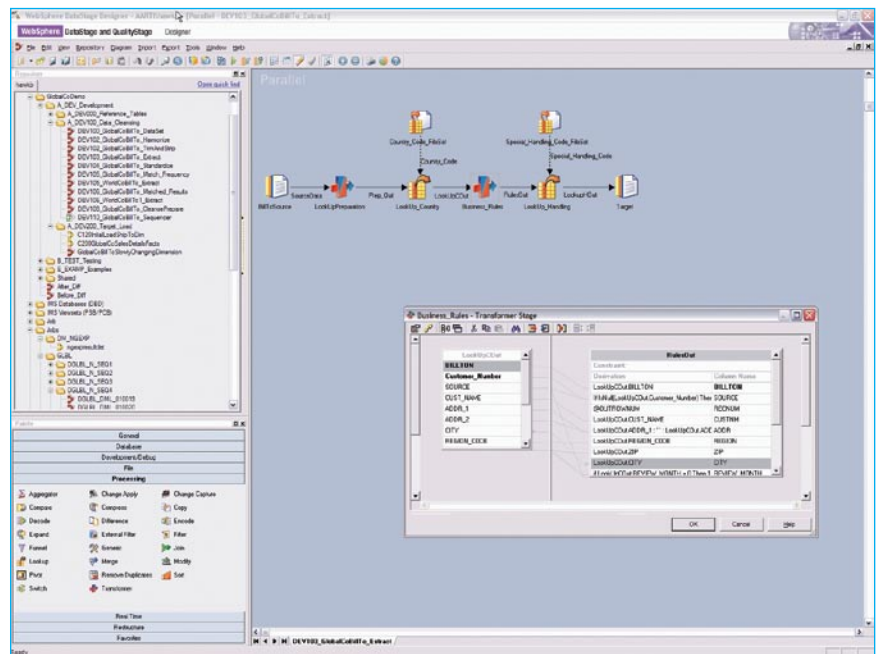

Figure 2: IBM WebSphere DataStage makes it easy to design enterprise data flows using a top-down, work-as-you-think metaphor.

The most flexible and scalable platform available

IBM Information Server enables companies to solve large-scale business problems by enabling high-performance integration of large data volumes. By leveraging the parallel processing capabilities of multiprocessor hardware platforms, IBM Information Server linearly increases the speed of data throughput. It can scale transformation jobs to satisfy the demands of ever-growing data volumes and ever-shrinking batch windows.

Development is done using sequential logic and the deployment configuration automatically adds the desired degree of parallelism. As a result, should a business need to boost the frequency of its integration, users could take the application from 2-way processing in the morning to 32-way in the afternoon to 128-way processing at night—all with only a simple change to the configuration file.

The secret: Partitioning and dynamic repartitioning

The IBM Information Server parallel technology operates using a divide-and-conquer technique, splitting the largest integration jobs into subsets (partition parallelism) and flowing these subsets concurrently across all available processors (pipeline parallelism). This combination of pipeline and partition parallelism delivers true linear scalability—defined as an increase in performance proportional to the number of processors—and makes hardware the only mitigating factor to performance.

Consider a transformation based on customer last name, where the enriching needs to occur on zip code—for householding purposes—and credit card number for loading into the data warehouse database parallel loader. With dynamic data repartitioning, data is repartitioned on the fly between processes—without landing the data to disk, a slow and costly step required by many other integration products.

Wide-ranging parallel support for SMP, MPP and grid deployments

IBM Information Server scales effortlessly from symmetric multiprocessor (SMP) and SMP clusters to massively parallel processing (MPP) servers with hundreds of processors. The same integration capability is available on a grid deployment of low-cost servers. This wide-ranging support for parallel

processing helps ensure critical enterprise information integration applications will scale in pace with business requirements.

IBM Information Server transformation features

- *Easy-to-use, top-down, work-as-you-think design interface*
- *Service Oriented Architecture (SOA) for creating data integration and transformation services for real-time deployment and for making trusted information available to a larger development community and set of applications without having to re-create complex integration data flows*
- *Common metadata repository that helps ensure seamless integration with other IBM Information Server product modules including Data Profiling and Data Quality*

- *The most comprehensive library of transformation components for easily defining common integration processes and an architecture to support sharing and reuse*
- *Powerful tools to administer, deploy and update the data flows throughout the entire data integration life cycle*
- *Industry-leading connectivity to operational systems, databases and enterprise applications spread across mainframe and distributed systems*
- *Parallel processing support that enables users to design once and deploy as needed at run time without having to change any of the integration jobs*

Transformation within a unified platform

IBM WebSphere DataStage® is the core product module of IBM Information Server that delivers data integration and transformation capabilities for unsurpassed levels of productivity.

WebSphere DataStage is available on the following platforms:

- *Microsoft Windows Server® 2003*
- *IBM AIX®*
- *HP-UX*
- *Sun Solaris*
- *Red Hat Enterprise Linux® AS*
- *SUSE Linux Enterprise*

For detailed information about transforming data natively on the IBM z/OS® platform, please see the IBM documents *IBM WebSphere DataStage for z/OS* and *IBM WebSphere DataStage MVS™ Edition.*

National language support

WebSphere DataStage is National Language Support–enabled using Unicode.

IBM Information Server delivers trusted information

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 Information Server is a revolutionary new software platform that helps organizations derive more value from the complex, heterogeneous information spread across their 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 to collaborate to understand the meaning, structure and content of any type of information across any sources. It provides break-through 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 Information Server, contact your IBM marketing representative or IBM Business Partner, or visit **ibm.com**/software/data/integration

IBM®

TAKE BACK CONTROL WITH **Information Management**

IMD11783-USEN-00