

IBM WebSphere DataStage Version 7.5

Highlights

- ***Access, transform and deliver information from multiple sources, in multiple formats, regardless of the volume of data or time frame required for processing***
- ***Quickly deliver integrated views of enterprise data without complex, hand coded integration processes***
- ***Enable consistent application of business rules across data integration projects***
- ***Reduce the costs and risks associated with complex business initiatives***

IBM WebSphere Information Integration

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 list goes on, and the challenges grow unceasingly if businesses cannot ensure that they have access to authoritative, consistent, timely and complete information.

The IBM WebSphere® Information Integration platform integrates and transforms any data and content to deliver information you can trust for your critical business initiatives. It provides breakthrough productivity, flexibility and performance, so you and your customers and partners have the right information for running and growing your businesses. It helps you understand, cleanse and enhance information, while governing its quality to ultimately provide authoritative information. Integrated across the extended enterprise and delivered when you need it, this consistent, timely and complete information can enrich business processes, enable key contextual insights and inspire confident business decision-making.

IBM WebSphere DataStage®, a core component of the IBM WebSphere Data Integration Suite, enables you to tightly integrate enterprise information, despite having many sources or targets and short time frames. Whether you're 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 customer relationship management (CRM), supply chain management (SCM), and enterprise resource planning (ERP), WebSphere DataStage helps ensure that you will have information you can trust.

The industry's most powerful solution

WebSphere DataStage delivers three key capabilities necessary for success in enterprise data integration: the most 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 volumes of corporate data.

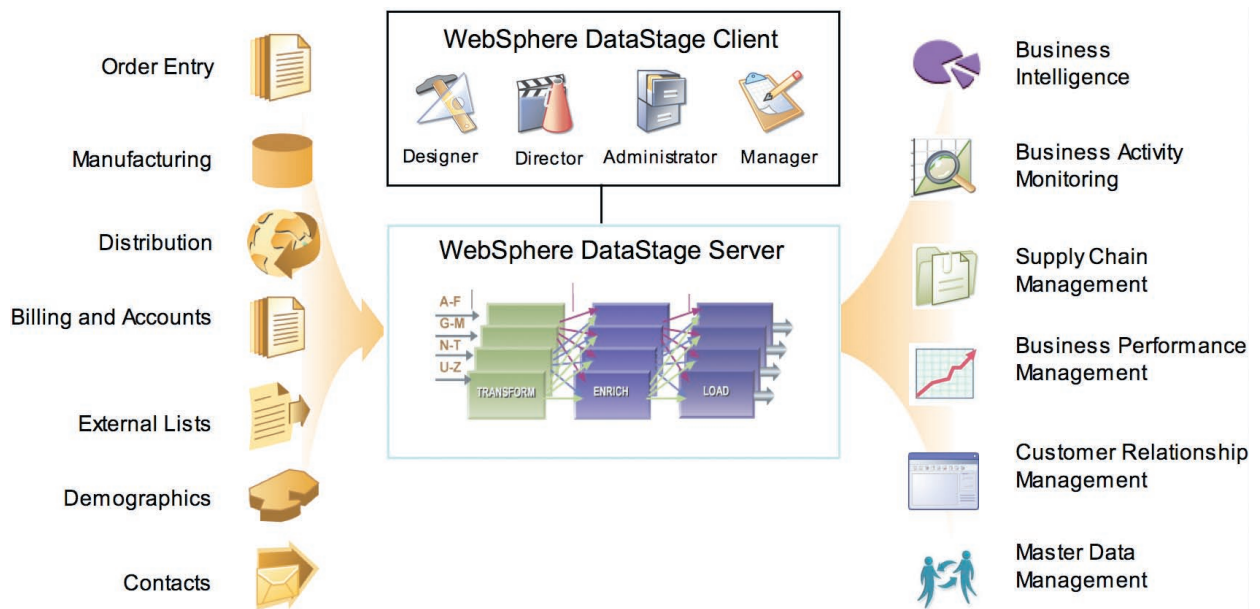


Figure 1. Connect, Extract, Transform and Deliver

WebSphere DataStage supports the collection, integration and transformation of high volumes of data, with data structures ranging from simple to highly complex. WebSphere DataStage manages data arriving within seconds of being acquired, as well as massive quantities of data that flood the system, in daily, weekly or monthly processing intervals.

The most comprehensive connectivity to source and target systems

WebSphere DataStage supports a virtually unlimited number of heterogeneous data sources and targets in a single job, including:

- *Text files*
- *Complex XML data structures*
- *Enterprise application systems including 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 products including WebSphere MQ and SeeBeyond*

And the list goes on. If it's in your enterprise, it's supported. WebSphere DataStage's comprehensive connectivity to enterprise data sources means you no longer need to develop and maintain custom code to connect to your important data.

Right time data integration

WebSphere DataStage can operate in real time, capturing messages or extracting data at a moment's notice on the same platform that also integrates bulk data. This provides a key advantage over competing offerings that require the use of two separate tools to achieve the same functionality, and allows you to respond to your organization's data integration needs on demand.

Advanced development and maintenance

WebSphere DataStage features 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 in WebSphere DataStage reduce learning curves, simplify administration, and optimize the use of development resources resulting in a decreased development and maintenance cycle for data integration applications. As a result, WebSphere DataStage enables companies to spend less time developing their integration and more time reaping the benefits of it.

Complete development environment

WebSphere DataStage employs a “work as you think” design metaphor. Developers use a top down dataflow model of application programming and execution, which allows them to create a visual sequential data flow. A robust graphical palette helps developers diagram the flow of data through their environment via simple GUI-driven drag-and-drop design components. Developers also benefit from a versatile scripting language, powerful debugging capabilities, and an open application programming interface (API) for leveraging external code.

Get started quickly

WebSphere DataStage includes Intelligent Assistants to make frequently used tasks easy to create. Intelligent Assistants are wizard-like functionality used for initial job creation. Job templates and preconfigured components also speed development.

Powerful pre-built functions

WebSphere DataStage features the industry's most extensive data integration development environment, with a library of more than 400 pre-built functions and routines. This complete set of data transformation capabilities makes it easy to map data from source to target and enrich it along the way.

Reuse, versioning and sharing

WebSphere DataStage shortens the development cycle by promoting the reuse of existing data integration business logic. This works through the concept of containers, which allow jobs and meta data created in one container to be shared and reused by other jobs. Versioning extends the development, test and deployment of jobs among multiple developers or WebSphere DataStage servers. Robust job specification reporting provides documentation so other developers can easily understand job design and provide additional support.

The most scalable platform available

WebSphere DataStage enables companies to solve large-scale business problems through high-performance processing of massive data volumes. By leveraging the parallel processing capabilities of multi-processor hardware platforms, IBM WebSphere DataStage Enterprise Edition can scale to satisfy the demands of ever growing data volumes and ever shrinking batch windows.

WebSphere DataStage Enterprise Edition minimizes the time-processing requirements and, by fully leveraging the parallel processing capabilities, linearly increases speed of data throughput for integrating massive amounts of data. It also boosts developer productivity by eliminating the need to code new applications to run in parallel—a costly process that often requires the expertise of specialists. Development is done using sequential logic and the deployment configuration automatically adds the desired degree of parallelism.

Open and extensible

WebSphere DataStage Enterprise Edition is a robust, open environment that not only supports IBM WebSphere integration products like IBM WebSphere ProfileStage™ and IBM WebSphere QualityStage™, but also third-party applications like SAS. In addition, WebSphere DataStage supports custom, homegrown code, enabling companies to reuse their existing proprietary code and execute it in parallel against unlimited data volumes.

Flexible parallelism

A separate configuration file allows users to define the degree of parallelism without changes to application code. As a result, should the business need to boost the frequency of its integration, users could take the application from two-way 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 re-partitioning

IBM's parallel technology operates by 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. However, downstream processes may need data partitioned differently. Consider a transformation that is based on customer last name, with the enriching needs to occur on zip code—for house-holding purposes—with loading into the warehouse based on customer credit card number (more on parallel database interfaces below). With dynamic data re-partitioning, data is re-partitioned on-the-fly between processes—without landing the data to disk—based on the downstream process data partitioning needs.

Wide-ranging parallel hardware support

WebSphere DataStage scales effortlessly from symmetric multiprocessor (SMP) and SMP clusters to massively parallel processing (MPP) servers with hundreds of processors. This capability ensures critical enterprise information integration applications will scale in pace with business.

Products

WebSphere DataStage

WebSphere DataStage is the industry-leading data integration and transformation product that provides advanced development and maintenance capabilities for unsurpassed levels of productivity.

WebSphere DataStage Extended Edition

IBM WebSphere DataStage Extended Edition builds upon WebSphere DataStage by incorporating IBM WebSphere MetaStage® meta data management solution for a clear, unambiguous definition and history of your data, and the WebSphere DataStage Pack for Web Services, which enables WebSphere DataStage designers to leverage web services-based resources to enrich their job design, or as source and target information remotely. Message adapters such as IBM WebSphere MQ, are also included with WebSphere DataStage Extended Edition.

WebSphere DataStage

Enterprise Edition

IBM WebSphere DataStage Enterprise Edition takes performance to a new level. Parallel processing capabilities, including partitioning, dynamic re-partitioning, parallel database interfaces, and exploitation of scalable hardware environments allows you to handle the massive volume, velocity and variety of data flowing into your organization.

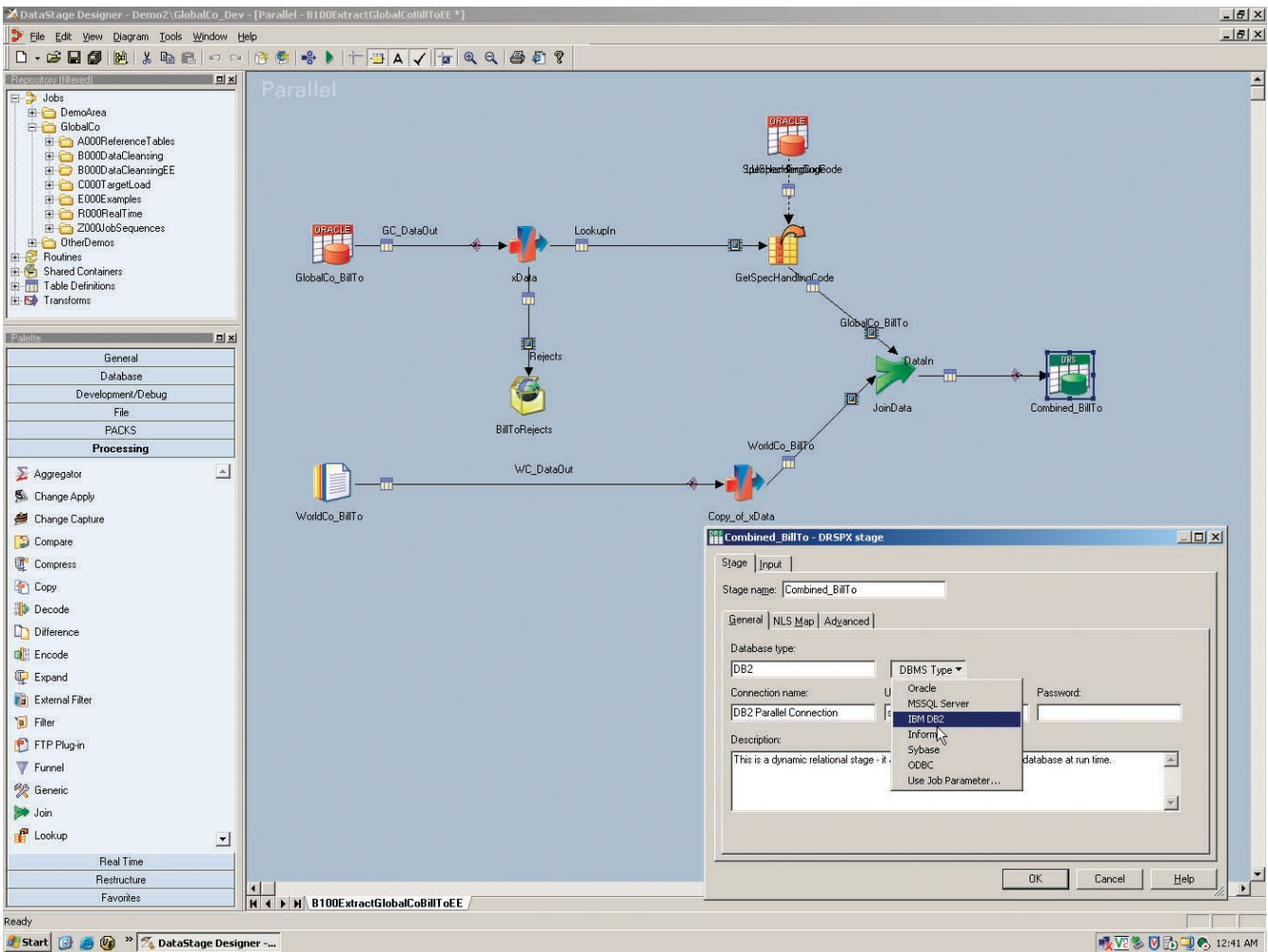


Figure 2. WebSphere DataStage Transformer

Together with end-to-end meta data management, advanced maintenance and development, and the ability to operate in real-time, WebSphere DataStage Enterprise Edition provides the most powerful data integration and transformation solution available.

WebSphere DataStage for z/OS

IBM WebSphere DataStage for z/OS® supports Unix System Services (USS) on IBM z/OS servers. This capability allows you to use the data where it resides—on the mainframe. By taking advantage of the USS system’s high utilization rates and WebSphere DataStage’s parallel capabilities, customers can gain more leverage to accommodate shrinking batch windows.

WebSphere DataStage MVS Edition

IBM WebSphere DataStage MVS™ Edition enables customers to leverage the power of mainframe resources to process data where it resides. This edition uses the WebSphere DataStage design tools to generate COBOL programs and JCL scripts that execute natively on the mainframe.

Technical specifications

	WebSphere DataStage	WebSphere DataStage Extended Edition	WebSphere DataStage Enterprise Edition
Platforms	Windows 2000 Windows Server 2003 IBM AIX® HP Compaq Tru64 HP HP -UX Sun Solaris Red Hat Enterprise Linux AS SuSE Enterprise Linux	IBM AIX HP HP -UX Sun Solaris Red Hat Enterprise Linux AS SuSE Enterprise Linux	IBM AIX HP HP -UX Sun Solaris Red Hat Enterprise Linux AS SuSE Enterprise Linux
WebSphere MetaStage	Available	Included	Included
Web Services Client Pack	Available	Included	Included
Adapters	Available	Included	Included

For current, detailed hardware and software system requirements for these and other WebSphere Information Integration products, visit ibm.com/software/data/integration.

WebSphere DataStage for z/OS

Development is done using a Windows-based DataStage Client connected to a Unix WebSphere DataStage Server.

Job compilation and execution are done using either a connected tightly coupled model or a loosely coupled mechanism.

In the tightly coupled mode:

- *Jobs can be designed, compiled, and run under the control of the WebSphere DataStage Clients*
- *Login and monitoring information is available in WebSphere DataStage*
- *Remote shell (rsh) and FTP are used automatically to connect to the mainframe*

In the loosely coupled mode:

- *The job scripts can be either automatically sent to the mainframe using FTP, or sent manually*
- *All jobs may be executed using command-line interfaces or a mainframe scheduler*
- *Job logging and monitoring information is not returned to the WebSphere DataStage Server if in this mode*

WebSphere DataStage MVS Edition

Development is done using a Windows-based WebSphere DataStage Client connected to a Unix WebSphere DataStage Server.

The MVS Edition generates COBOL applications and the corresponding custom JCL scripts for processing mainframe:

- *Flat files*
- *IBM DB2*
- *IBM IMS*
- *Teradata*

Users can also integrate custom in-house applications into the design.

National language support

WebSphere DataStage is
National Language Support
(NLS) enabled using Unicode.

For more information

To learn more about the technologies
and products behind the IBM
WebSphere Information Integration
platform, contact your IBM
marketing representative or
IBM Business Partner, or visit
ibm.com/software/data/integration.



© Copyright IBM Corporation 2005
IBM Software Group
Route 100
Somers, NY 10589
U.S.A.

Printed in the United States of America
11-05
All Rights Reserved

AIX, DataStage, DB2 Universal Database, IBM, the IBM logo, IMS, Informix, MetaStage, MVS, the On Demand Business logo, ProfileStage, QualityStage, WebSphere and z/OS are trademarks of International Business Machines Corporation in the United States, other countries or both.

Linux is a registered trademark of Linus Torvalds in the United States, other countries or both.

Microsoft and Windows are registered trademarks of Microsoft Corporation in the United States, other countries or both.

UNIX is a registered trademark of The Open Group in the United States and other countries.
Other company, product 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. Offerings are subject to change, extension or withdrawal without notice.

Other company, product or service names may be trademarks or service marks of others.

All statements regarding IBM future direction or intent are subject to change or withdrawal without notice and represent goals and objectives only.

The IBM home page on the Internet can be found at **ibm.com**