



Highlights

- Scales for data of any size, regardless of volume and complexity
 - Provides agile, reusable integration across diverse sources
 - Helps users quickly respond to business changes
 - Provides the ability to run natively in Apache Hadoop for fast processing and enhanced scalability
 - Enables self-service data integration for traditional and big data projects
 - Available on the Cloud—Move existing workloads and accelerate the time to value of new deployments
 - Delivers integration for cloud environments
 - Provides enriched security
-

Flexible integration with IBM InfoSphere DataStage V11.5

Built on a massively parallel processing (MPP) architecture, IBM® InfoSphere® DataStage® V11.5 is designed to help organizations transform and integrate large volumes of heterogeneous data. InfoSphere DataStage enables users to design jobs once and deploy anywhere, resulting in improved performance, greater integration agility and lower costs.

A new era of computing is unfolding, bringing with it an ever-accelerating explosion in the volume, variety and velocity of data. To achieve their business intelligence goals and gain a competitive edge, organizations must be able to integrate data from anywhere in their environment and transform it into trusted information. The ability to integrate information quickly and efficiently is crucial in this dynamic environment, even as those requirements continue to shift and change and data volumes increase.

The end-to-end information capabilities of InfoSphere Information Server allow you to understand your data, cleanse, monitor, transform and deliver it, as well as collaborate to bridge the gap between business and IT. By leveraging the “anywhere integration” capabilities delivered by InfoSphere Information Server, you can ensure the information that drives your business and your strategic initiatives—from big data and point-of-impact analytics to master data management (MDM) and data warehousing—is trusted, consistent and governed in real time.



DataStage features address complex data integration challenges

InfoSphere DataStage V11.5 offers powerful new capabilities for today's information-rich environments.

Easy scalability and high performance for fast access to clean data

A massively parallel processing engine that runs natively in Hadoop where the data resides enables a uniquely rich set of integration and governance features. These features help your organization achieve high levels of performance and speed for data integration and governance activities, such as data connectivity, transformation, cleansing, enrichment and delivery features. For example:

- All connectivity, transformation and data delivery features can now run natively in Hadoop
- Provides expanded and simplified access to Hadoop Distributed File System (HDFS) files in various formats and character sets, including critical security features, such as Kerberos and secure gateways

Lower cost of integration due to increased productivity

The InfoSphere Information Server engine gives your developers access to built-in graphical design tool capabilities. Based on our customers' experiences, this will help your organization significantly reduce the cost of integration compared to manual hand-coding.

Available in the Cloud

A new way to work—flexibility for hybrid environments integrating all types of data

Whether you want to move your entire InfoSphere DataStage infrastructure to the cloud or you want to start small by deploying development or test environments IBM DataStage

on Cloud is ready to help with your data integration needs through rapid deployment and flexible subscription pricing.

If you are brand new to IBM DataStage you can begin on the cloud and avoid the steps to procure on-premises hardware, departmental and budget approvals with low-risk subscription pricing for DataStage on Cloud. Rapidly deploy the entire ETL infrastructure that your business needs.

Additional features at no extra cost

Your organization can benefit from additional features such as access to popular enterprise applications, including:

- Salesforce.com
- Hyperion Essbase
- Siebel
- JD Edwards EnterpriseOne
- PeopleSoft Enterprise
- Oracle applications

You can also take advantage of push-down optimization through InfoSphere DataStage Balanced Optimization features, which allow you to fully harness available capacity and computing power in your relational databases and in InfoSphere DataStage.

Data integration for cloud environments

InfoSphere DataStage V11.5 provides quick and easy data integration for cloud environments. When running on premises, it supports direct integration with the Amazon Simple Storage System (S3) to load data from and into the cloud. Once data is integrated within S3, it may be then be integrated with other cloud database technologies.

Additionally, InfoSphere DataStage V11.5 includes a Hierarchical Stage (renamed and expanded from the "XML Stage") that supports interaction with REST application programming interfaces (APIs), enabling integration support for XML and JavaScript Object Notation (JSON) messages.

The REST-based connectivity enables InfoSphere Information Server to support distributed database-as-a-service (DBaaS) offerings such as IBM Cloudant®, as well as other on-premises and off-premises solutions that offer REST-based interaction.

Tight integration for master data management projects

InfoSphere DataStage can directly integrate with InfoSphere MDM through an MDM Integration Stage, enabling users to easily employ InfoSphere DataStage to load data into and extract data out of InfoSphere MDM. By leveraging the MDM Integration Stage, customers can include MDM data within their data integration flows and can load domain data (such as customer, partner, supplier, product and other data) directly to an MDM system. Additionally, users may now benefit from peak scale and performance by leveraging bulk loading to achieve the best performance.

In addition to benefiting from the InfoSphere Information Server data transformation and delivery styles, the MDM Integration Stage can leverage InfoSphere Information Server data quality capabilities as well. For example, InfoSphere Information Server can standardize any data before it is loaded into the MDM system. By proactively addressing data quality in this way, projects benefit from improved matching accuracy and better support for 360-degree views of various entities.

Big data integration: A key to big data success

A fast and easy way to get data into and out of big data distributions is a must-have in today's fast-moving business climate. Thanks to an MPP architecture, InfoSphere DataStage can easily scale to meet the demanding data integration requirements of Hadoop environments. The InfoSphere DataStage “design once, run anywhere” approach allows users to easily

move data integration tasks between a single machine and a cluster of low-commodity servers. Because job design is isolated from runtime, developers can focus on the business requirements at hand rather than coding explicitly for a given architecture.

InfoSphere DataStage always has been MPP-based, and it is the original integration platform to support extreme (big) data volumes—even before the term “big data” came into vogue. InfoSphere DataStage delivers powerful big data-specific capabilities to help your big data projects proceed quickly and cost-efficiently. These include:

- **Balanced optimization for Hadoop:** This feature allows integration designers to build a job using the same design paradigm they use with traditional extract, transform and load (ETL) and, when desired, run that logic through generated MapReduce. This eliminates the cost of retooling and training the team on MapReduce languages or other secondary toolsets that would apply only to this particular environment.
- **Big data job sequencing:** InfoSphere DataStage allows any Oozie-contained MapReduce job to be included in the job sequencer. Organizations can build workflows that load data to Hadoop, run a custom-developed MapReduce analytics program, and then load the analytical result to the data warehouse—all within a single graphical workflow construct. This provides end-to-end workflow across heterogeneous topologies executed in both InfoSphere Information Server and Hadoop.
- **Big data governance:** InfoSphere DataStage supports big data-related governance features such as impact analysis and data lineage on any integration points, enabling scalable analytics without sacrificing organizational insight.

- **IBM InfoSphere Streams integration:** For big data projects that focus on real-time analytical processing, IBM now offers direct data flow integration between InfoSphere DataStage and InfoSphere Streams. Organizations can use standard data integration conventions to gather and pass information to real-time analytical processes.
- **Big data accelerators:** These open source components are available on ibm.com/developerworks. They plug directly into the InfoSphere DataStage canvas and operate just like any other stage. Accelerators are available for MongoDB, Hive, Cassandra, HBase, Avro and more.
- **Expanded hierarchical data support:** Version 11.5 expands InfoSphere DataStage support to provide REST API support, allowing easy access to and integration of hierarchical data, such as XML and JSON messages.

Agile, self-service data provisioning and governance

InfoSphere Data Click, available as part of InfoSphere DataStage V11.5, helps speed up time to value, increase business agility and lower costs by shrinking the time required to complete tasks from days or weeks to minutes or hours. It provides broad connectivity, helping to improve the timeliness of many different types of environments, including big data landing zones, data warehouses and cloud environments. InfoSphere Data Click provides both specific native DBMS connectivity for fast performance and near-universal connectivity for a very wide range of data sources through JDBC and ODBC (Figure 1).

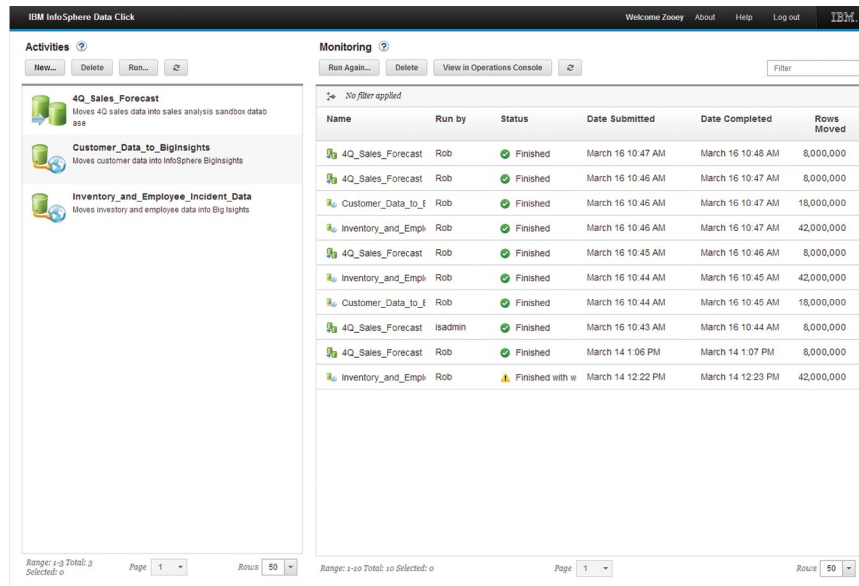


Figure 1. Users log directly into a streamlined InfoSphere Data Click UI. From here, they can quickly and easily create, edit, run and monitor activities.

Data can be sourced from or sent to a whole host of traditional and big data environments, including IBM DB2®, Oracle, EMC Greenplum, IBM Informix®, IBM PureData® for Analytics (based on IBM Netezza®), Salesforce.com, Microsoft SQL Server, Teradata, Amazon S3 and more. InfoSphere Data Click also allows any user to easily provision data within their IBM BigInsights® (that is, Hadoop) environment.

Enhanced security for InfoSphere Information Server

InfoSphere Information Server delivers common platform services (such as connectivity services, administration services, deployment services and so on) to support its data integration, data quality and data governance capabilities. Along with cloud enhancements and additional support for massive data volumes, V11.5 includes:

- **Single sign-on (SSO):** Browser-based clients now support SSO, allowing customers who authenticate within one of the interfaces to more seamlessly interact across other user interfaces.
- **Secure Sockets Layer (SSL) communication:** InfoSphere Information Server employs SSL to provide communication security for all client interfaces.
- **Stronger encryption:** RSA-2048 and SHA-512 are available as default encryption mechanisms.
- **Cell sharing:** InfoSphere Information Server fully leverages the IBM WebSphere® Application Server standard security domain. As such, it can be deployed into an existing cell managed by a secured deployment manager without disrupting the profiles and applications already deployed in that cell.

A step further: InfoSphere Information Server for Data Integration

Together, InfoSphere DataStage and InfoSphere Information Server enable organizations to flexibly and robustly manage big data from new and emerging sources.

InfoSphere DataStage users with broader data integration requirements may also be interested in InfoSphere Information Server for Data Integration V11.5. This solution includes InfoSphere DataStage, along with additional, frequently requested data integration capabilities such as change data delivery, real-time data integration, data modeling, blueprinting, data discovery and governed metadata management. It also contains functionality to accelerate design time, creating source-to-target mappings and automatically generating jobs.

To meet other information integration requirements, IBM also offers InfoSphere Information Server for Data Quality, InfoSphere Information Governance Catalog and InfoSphere Information Server Enterprise Edition. The last of these packages is a comprehensive offering that includes all three sets of capabilities—data integration, data quality and information governance.

Why IBM?

As a critical element of IBM Watson™ Foundations, InfoSphere Information Integration and Governance (IIG) provides market-leading functionality to handle the challenges of big data. InfoSphere IIG provides optimal scalability and performance for massive data volumes, agile and right-sized integration and governance for the increasing velocity of data, and support and protection for a wide variety of data types and big data systems whether on-premises or in the cloud. InfoSphere IIG helps make big data and analytics projects successful by giving business users the confidence to act on insight.

InfoSphere capabilities include:

- **Metadata, business glossary and policy management:** Define metadata, business terminology and governance policies with InfoSphere Information Governance Catalog.
- **Data integration:** Handle all integration requirements, including batch data transformation and movement (InfoSphere Information Server), real-time replication (InfoSphere Data Replication) and data federation (InfoSphere Federation Server).
- **Data quality:** Parse, standardize, validate and match enterprise data with InfoSphere Information Server for Data Quality.
- **MDM:** Act on a trusted view of your customers, products, suppliers, locations and accounts with InfoSphere MDM.
- **Data lifecycle management:** Manage the data lifecycle from test data creation through retirement and archiving with IBM InfoSphere Optim™.
- **Data Privacy:** Help ensure sensitive data is masked and protected with InfoSphere Optim.

For more information

To learn more about InfoSphere Information Server, InfoSphere Information Server for Data Integration or InfoSphere DataStage, or DataStage on Cloud, contact your IBM representative or IBM Business Partner, or visit the following websites:

- ibm.com/software/products/en/ibminfodata
- ibm.com/software/products/en/infoinfoservfordatainte

Additionally, IBM Global Financing provides numerous payment options to help you acquire the technology you need to grow your business. We provide full lifecycle management of IT products and services, from acquisition to disposition. For more information, visit: ibm.com/financing



© Copyright IBM Corporation 2016

IBM Analytics
Route 100
Somers, NY 10589

Produced in the United States of America
June 2016

IBM, the IBM logo, ibm.com, BigInsights, Cloudant, DataStage, DB2, IBM Watson, Informix, InfoSphere, Optim, PureData, and WebSphere 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

Netezza is a trademark or registered trademark of IBM International Group B.V., an IBM Company.

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

Java and all Java-based trademarks and logos are trademarks or registered trademarks of Oracle and/or its affiliates.

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